Review of Economic Studies (2011) **78**, 356–393 doi: 10.1093/restud/rdq014

Social Learning and Peer Effects in Consumption: Evidence from Movie Sales

ENRICO MORETTI

University of California, Berkeley, NBER, CEPR, and IZA

First version received August 2008*; final version accepted January* 2010 (*Eds.*)

Using box-office data for all movies released between 1982 and 2000, I quantify how much the consumption decisions of individuals depend on information they receive from their peers, when quality is *ex ante* uncertain. In the presence of social learning, we should see different box-office sales dynamics depending on whether opening weekend demand is higher or lower than expected. I use a unique feature of the movie industry to identify *ex ante* demand expectations: the number of screens dedicated to a movie in its opening weekend reflects the sales expectations held by profit-maximizing theatre owners. Several pieces of evidence are consistent with social learning. First, sales of movies with positive surprise and negative surprise in opening weekend demand diverge over time. If a movie has better than expected appeal and therefore experiences larger than expected sales in Week 1, consumers in Week 2 update upward their expectations of quality, further increasing Week 2 sales. Second, this divergence is small for movies for which consumers have strong priors and large for movies for which consumers have weak priors. Third, the effect of a surprise is stronger for audiences with large social networks. Finally, consumers do not respond to surprises in first-week sales that are orthogonal to movie quality, like weather shocks. Overall, social learning appears to be an important determinant of sales in the movie industry, accounting for 32% of sales for the typical movie with positive surprise. This implies the existence of a large "social multiplier" such that the elasticity of aggregate demand to movie quality is larger than the elasticity of individual demand.

Keywords: Social interactions, Social multiplier, Peer effects

JEL Codes: L0, J0

1. INTRODUCTION

The objective of this paper is to estimate how much the consumption decisions of individuals depend on information they receive from their peers when product quality is difficult to observe in advance. I focus on situations where quality is *ex ante* uncertain and consumers hold a prior on quality, which they may update based on information from their peers. This information may come from direct communication with peers who have already consumed the good. Alternatively, it may arise from the observation of peers' purchasing decisions. If every individual receives an independent signal of the goods quality, then the purchasing decision of one consumer provides valuable information to other consumers as individuals use the information contained in others' actions to update their own expectations on quality.

This type of social learning is potentially relevant for many experience goods like movies, books, restaurants, or legal services. Informational cascades are particularly important for new products. For the first few years of its existence, Google experienced exponential acceleration in market share. This acceleration, which displayed hallmarks of contagion dynamics, was mostly due to word of mouth and occurred without any advertising on the part of Google (Vise, 2005).

Social learning in consumption has enormous implications for firms. In the presence of informational cascades, the return to attracting a new customer is different from the direct effect that the customer has on profits. Attracting a new consumer has a multiplier effect on profits because it may increase the demand of other consumers. The existence of this "social multiplier" (Glaeser, Sacerdote and Scheinkman, 2003) implies that, for a given good, the elasticity of aggregate demand to quality is larger than the elasticity of individual demand to quality. Furthermore, social learning makes the success of a product more difficult to predict as demand depends on (potentially random) initial conditions. Two products of similar quality may have vastly different demand in the long run, depending on whether the initial set of potential consumers happens to like the product or not.

Social learning has been extensively studied in theory (Banerjee, 1992; Bikhchandani, Hirshleifer and Ivo, 1992, 1998). But despite its tremendous importance for firms, the empirical evidence is limited because social learning is difficult to identify in practice. The standard approach in the literature on peer effects and social interactions involves testing whether an individual decision to purchase a particular good depends on the consumption decisions and/or the product satisfaction of other individuals that are close, based on some metric. Such an approach is difficult to implement in most cases. First, data on purchases of specific goods by individual consumers are difficult to obtain. Second, because preferences are likely to be correlated among peers, observing that individuals in the same group make similar consumption decisions may simply reflect shared preferences not informational spillovers. In observational data, it is difficult to isolate factors that affect some individuals' demand for a good but not the demand of their peers.[1]

In this paper, I focus on consumption of movies. Since individual-level data are not available, I use market-level data to test the predictions of a simple model that characterizes the diffusion of information on movie quality following *surprises* in quality. I assume that the quality of a movie is *ex ante* uncertain as consumers do not know for certain whether they will like the movie or not.[2] Consumers have a prior on quality—based on observable characteristics of the movie such as the genre, actors, director, ratings and budget, etc.—and they receive an individual-specific, unbiased signal on quality—which reflects how much the concept of a movie resonates with a specific consumer.

I define social learning as a situation where consumers in week t update their prior based on feedback from others who have seen the movie in previous weeks. The model predicts different box-office sales dynamics depending on whether a movie's underlying quality is better or worse than people's expectations. Because the signal that each consumer receives is unbiased, movies that have better than expected underlying quality have stronger than expected demand in the opening weekend (on average). In the presence of social learning, they become even more successful over time as people update upward their expectations on quality. On the other hand, movies that have worse than expected quality have weaker than expected demand in the opening weekend (on average) and become even less successful over time. In other words, social learning should make successful movies more successful and unsuccessful movies more unsuccessful. By contrast, without social learning, there is no updating of individual expectations, and therefore there should be no divergence in sales over time.

1. Randomized experiments may offer a solution. Salganik, Dodds and Watts (2006) set up a website for music downloading where users are randomly provided with different amount of information on other users' ratings. Cai, Chen and Fang (2007) use a randomized experiment to study learning on menu items in restaurants.

2. Throughout the paper, the term quality refers to consumers' utility. It has no reference to artistic value.

Surprises in the appeal of a movie are key to the empirical identification. I use a unique feature of the movie industry to identify *ex ante* demand expectations: the number of screens dedicated to a movie in its opening weekend reflects the sales expectations held by the market. The number of screens is a good summary measure of *ex ante* demand expectations because it is set by forward-looking, profit-maximizing agents—the theatre owners—who have an incentive to correctly predict first-week demand. The number of screens should therefore reflect most of the information that is available to the market before the opening on the expected appeal of the movie, including actors, director, budget, ratings, advertising, reviews, competitors, and every other demand shifter that is observed before opening day.

While on average theatres predict first-week demand correctly, there are cases where they underpredict or overpredict the appeal of a movie. Take, *e.g.* the movie "Pretty Woman" (1990). Before the opening weekend, it was expected to perform well, since it opened in 1325 screens, more than the average movie. But in the opening weekend, it significantly exceeded expectations, totalling sales of about $23 million. In this case, demand was significantly above what the market was expecting, presumably because the concept of the movie or the look of Julia Roberts appealed to consumers more than one could have predicted before the opening. I use a Mincer–Zarnowitz test to check whether empirically the estimated prediction error is orthogonal to features of the movie that are known to theatres on opening day, as implied by my assumption. The results of the test are not definitive. Taken individually, observables are not statistically different from zero at conventional levels, although taken jointly they have a p-value of $0{\cdot}10$.

Using data on nationwide sales by week for all movies released between 1982 and 2000, I test five empirical implications of the model.

(1) In the presence of social learning, sales trends for positive and negative surprise movies should diverge over time. Consistent with this hypothesis, I find that the decline over time of sales for movies with positive surprise is substantially slower than the decline of movies with negative surprise. This finding is robust to controlling for advertising expenditures and critic reviews and to a number of alternative specifications. Moreover, the finding does not appear to be driven by changes in supply or capacity constraints. For example, results are not sensitive to using per-screen sales as the dependent variable instead of sales or dropping movies that sell out in the opening weekend.

(2) The new information contained in peer feedback should be more important when consumers have more diffuse priors. When a consumer is less certain whether she is going to like a specific movie, the additional information represented by peer feedback on movie quality should have more of an effect on her purchasing choices relative to the case where the consumer is more certain. In practice, to identify movies for which consumers have more precise priors, I use a dummy for sequels. It is reasonable to expect that consumers have more precise priors for sequels than non-sequels. Additionally, to generalize this idea, I use the variance of the first-week surprise in box-office sales by genre. Genres with large variance in first-week surprise are characterized by more uncertainty and therefore consumers should have more diffuse priors on their quality. Consistent with social learning, I find that the impact of a surprise on subsequent sales is significantly smaller for sequels and significantly larger for genres that have a large variance in first-week surprise.

(3) Social learning should be stronger for consumers with a large social network and weaker for consumers with a small social network. While I do not have a direct measure of social network, I assume that teenagers have more developed social networks than adults. Consistent with social learning, I find that the effect of a surprise on subsequent sales is larger for movies that target teenage audiences.

(4) The marginal amount of learning should decline over time as more information on quality becomes available. For example, the amount of updating that takes place in Week 2 should be

larger than the amount of updating that takes place in Week 3 given what is already known in Week 2. Consistent with this prediction, I find that sales trends for positive surprise movies are concave and sales trends for negative surprise movies are convex.

(5) Under social learning, surprises in opening weekend demand should only matter insofar as they reflect new information on movie quality. They should not matter when they reflect factors that are unrelated to movie quality.

This prediction is important because it allows me to separate social learning from a leading alternative explanation of the evidence, namely network externalities. A network externality arises when a consumer's utility of watching a movie depends directly on the number of peers who have seen the movie.[3] In the context of movies, this could happen, *e.g.* because people like discussing movies with friends, so that the utility from watching a movie that others in the peer group have seen is higher than the utility of watching the same movie when no one else in the peer group has seen it. Network externalities may also arise in the presence of preferences for conformity, *i.e.* when there is a utility premium in consuming products that are similar to the ones consumed by the reference group. Fashions are an example. This possibility could be particularly relevant for certain subgroups—like teenagers—for whom social pressure and conformity are salient.

Network externalities represent a very different explanation of the evidence than social learning. The social learning model assumes that individuals care about others' actions only because they convey information about the quality of a product. By contrast, network externalities imply that each consumer's utility from a good depends *directly* on the consumption by others. In the extreme case where the quality of a product is perfectly known in advance, there is no scope for social learning, while network externalities may still arise.

To distinguish between network externalities and social learning, I test whether consumers respond to surprises in first-week sales that are orthogonal to movie quality. In particular, I use an instrumental variable strategy to isolate surprises in first-week sales that are caused by weather shocks. Bad weather lowers sales but is independent of movie quality (after controlling for time of the year). Under social learning, a negative (positive) surprise in first-week sales caused by bad (good) weather should have no effect on sales in the following weeks. Since weather is unrelated to movie quality, surprises due to weather provide no quality signal and therefore should not induce any updating. By contrast, under the network externalities hypothesis, a negative surprise in first-week demand for any reason, including bad weather, should lower sales in following weeks. For example, if a consumer draws utility from talking about a movie with friends, she cares about how many friends have seen a movie, irrespective of their reasons for choosing to see it. Empirically, I find no significant effect of surprises due to weather on later sales.

Overall, the five implications of the model seem remarkably consistent with the data. Taken individually, each piece of empirical evidence may not be sufficient to establish the existence of social learning. But taken together, the weight of the evidence supports the notion of social learning.

My estimates suggest that the amount of sales generated by social learning is substantial. A movie with stronger than expected demand has \$4·5 million in additional sales relative to the counterfactual where the quality of the movie is the same but consumers do not learn from each other. This amounts to 32% of total revenues. To put this in perspective, consider that the effect on sales of social learning for the typical movie appears to be about two-thirds as large as the effect of all TV advertising. From the point of view of the studios, this implies the existence of a

3. Becker (1991) proposes a model of network effects where the demand for a good by a person depends positively on the aggregate quantity demanded of the good. He hypothesizes that "the pleasure from some goods is greater when many people want to consume it".

large multiplier. For a good movie, the total effect on profits of attracting an additional consumer is significantly larger than the direct effect on profits because that consumer will increase her peers' demand for the same movie.

Besides the substantive findings specific to the movie industry, this paper seeks to make a broader methodological contribution. It demonstrates that it is possible to identify social interactions using aggregate data and intuitive comparative statics. In situations where individual-level, exogenous variation in peer group attributes is not available, this approach has the potential to provide a credible alternative for the identification of social interactions. Possible applications include books, restaurants, cars, financial products, and the adoption of new technologies.

The paper proceeds as follows. In Sections 2 and 3, I outline a simple theoretical model and its empirical implications. In Sections 4 and 5, I describe the data and the empirical identification of surprises. In Sections 6, 7, and 8, I describe my empirical findings and their economic magnitude. In Section 9, I discuss other potential applications.

2. A SIMPLE MODEL OF SOCIAL LEARNING

In this section, I outline a simple framework that describes the effect of social learning on sales. The idea—similar to the one adopted by Bikhchandani, Hirshleifer and Ivo (1992) and Banerjee (1992)—is very simple. Consumers do not know in advance how much they are going to like a movie. Before the opening weekend, consumers share a prior on the quality of the movie—based on its observable characteristics—and they receive a private, unbiased signal on quality, which reflects how much the concept of a movie resonates with a specific consumer. Expected utility from consumption is a weighted average of the prior and the signal (where the weight reflects the relative precision of the prior and the signal). Since the consumers' private signal is unbiased, high-quality movies have *on average* a stronger appeal and therefore stronger opening weekend sales, relative to low-quality movies.

In Week 2, consumers have more information since they receive feedback from their peers who have seen the movie in Week 1. I define social learning as the process by which individuals use feedback from their peers to update their own expectations of movie quality. In the presence of social learning, a consumers expectation of consumption utility is a weighted average of the prior, the signal, and peer feedback, where, as before, the weights reflect relative precisions. If a movie is better (worse) than expected, consumers in Week 2 update upward (downward) their expectations and therefore even more (less) consumers decide to see the movie.

This setting generates the prediction that under social learning a movie whose demand is unexpectedly strong (weak) in the opening weekend should do even better (worse) in the following weeks. Without social learning, there is no reason for this divergence over time. The setting also generates four additional comparative statics predictions that have to do with the precision of the prior, the size of the social network, the functional form of movie sales, and the role of surprises that are orthogonal to quality. The model is designed to be simple and to generate transparent and testable predictions to bring to the data. As in Bikhchandani, Hirshleifer and Ivo (1992), I take the timing of consumption as exogenous. I purposely do not attempt to model possible generalizations such as strategic behaviour or the value of waiting to obtain more information (option value).

2.1. *Sales dynamics with no social learning*

The utility that individual i obtains from watching movie j is

$$U_{ij} = \alpha_j^* + \nu_{ij}, \tag{1}$$

where α_j^* represents the quality of the movie for the average individual and $\nu_{ij} \sim N\left(0, \frac{1}{d}\right)$ represents how tastes of individual *i* for movie *j* differ from the tastes of the average individual.

I assume that α_j^* and ν_{ij} are unobserved. Given the characteristics of a movie that are observed prior to its release, individuals hold a prior on the quality of the movie. In particular, I assume that

$$\alpha_j^* \sim N\left(X_j'\beta, \frac{1}{m_j}\right), \tag{2}$$

where $X_j'\beta$ represents consumers' priors on how much they will like movie j. Specifically, the vector X_j includes the characteristics of movie j that are observable before its release, including its genre, budget, director, actors, ratings, distribution, date of release, advertising, etc., and m_j is the precision of the prior, which is allowed to vary across movies. The reason for differences in the precision of the prior is that the amount of information available to consumers may vary across movies. For example, if a movie is a sequel, consumers may have a tighter prior than if a movie is not a sequel.

Before the release of the movie, I assume that each individual also receives a noisy, idiosyncratic signal of the quality of the movie:

$$s_{ij} = U_{ij} + \varepsilon_{ij}. \tag{3}$$

I interpret this signal as a measure of how much the concept of movie j resonates with consumer i. I assume that the signal is unbiased within a given movie and is normally distributed with precision k_j:

$$\varepsilon_{ij} \sim N\left(0, \frac{1}{k_j}\right). \tag{4}$$

The assumption that the prior and signal are unbiased is important because it ensures that, while there is uncertainty for any given individual on the true quality of the movie, *on average* individuals make the correct decisions regarding each movie. I also assume that ν_{ij} and ε_{ij} are i.i.d. and independent of each other and independent of α_j^*; that X_j, β, m_j, k_j, and d are known to all the consumers; and that consumers do not share their private signal.

The normal learning model indicates that the expected utility of the representative consumer in the opening weekend is a weighted average of the prior ($X_j'\beta$) and the signal (s_{ij}), with weights that reflect the relative precision of the prior and the signal:

$$E_1[U_{ij} \mid X_j'\beta, s_{ij}] = \omega_j X_j'\beta + (1 - \omega_j)s_{ij}, \tag{5}$$

where $\omega_j = \frac{h_j}{h_j + k_j}$, $h_j = \frac{dm_j}{d + m_j}$, and the subscript 1 on the expectation operator indicates that the expectation is taken using the information available in Week 1. When the prior is more (less) precise than the signal, prospective consumers put more (less) weight on the prior.

I assume that a consumer decides to see a movie if her expected utility given what she knows about the quality of the movie is higher than the cost:

$$E_1[U_{ij} \mid X_j'\beta, s_{ij}] > q_{i1}. \tag{6}$$

The cost of going to the movie at time t, q_{it}, may vary because it includes the opportunity cost of time, which varies across individuals and over time. For example, going to the movies may be very costly for a given individual if it conflicts with a work dinner, but it may be very cheap for the same individual on a different night. Similarly, the opportunity cost of time may vary across individuals. I assume that $q_{it} = q + u_{it}$, where $u_{it} \sim N\left(0, \frac{1}{r}\right)$ and i.i.d.

The probability that individual i goes to see movie j in the opening weekend is

$$P_1 = \text{Prob}(E_1[U_{ij} \mid X'_j\beta, s_{ij}] > q_{i1}) = \Phi\left(\frac{(1-\omega_j)(\alpha^*_j - X'_j\beta) + X'_j\beta - q}{\sigma_{j1}}\right), \qquad (7)$$

where $\Phi()$ is the standard normal cumulative function.[4]

The term $(\alpha^*_j - X'_j\beta)$ measures the surprise. In particular, it measures the distance between the true quality of the movie, α^*_j, and consumers' prior, $X'_j\beta$. Compare two movies with the same prior but with different true quality α^*_j. Imagine, *e.g.* that the quality of movie a is higher than its prior ($\alpha^*_a > X'_a\beta$) and the opposite is true for movie b ($\alpha^*_b < X'_b\beta$). Equation (7) indicates that movie a will experience higher opening weekend sales than the movie b. The reason is that the private signal received by consumers is unbiased, so that on average consumers find the better movie more attractive. If a movie sells well in the opening week, it is because the movie is of good quality and therefore many people received a good signal.

What happens to the number of tickets sold in the weeks after the opening weekend? In the absence of social learning, consumers in Week 2 and later weeks have exactly the same information that consumers have in Week 1. This implies that $P_t = P_1$, for $t \geq 2$. Therefore, a movie that does surprisingly well in the first-week experiences sales above its expected value, but the difference between actual sales and expected sales remains constant over time. This is represented as a parallel shift upward in the top panel of Figure 1. The amount of sales each week is higher, but the two lines are parallel. In the case of a movie that does surprisingly poorly in the first weekend, there is a parallel shift downward.

Note that sales are constant over time because I am implicitly assuming that consumers may watch the same movie multiple times. This assumption greatly simplifies the analysis. Below I show that the qualitative implications of the model do not change when consumers are assumed to see a movie only once, so that sales trend down over time.

2.2. *Sales dynamics with social learning*

With social learning, consumers have more information in Week 2 than they did in Week 1 because they receive feedback from their peers. Specifically, I assume that consumer i has N_i peers. Of these N_i peers, n_i see the movie in Week 1 and communicate to consumer i their *ex post* utility. Consumer i aggregates these feedbacks to obtain an unbiased estimate of quality and update her expected utility. I call this estimate S_{ij2}. In Appendix 1, I describe formally how the consumer obtains S_{ij2} and I show that $S_{ij2} \sim N\left(\alpha^*_j, \frac{1}{b_{i2}}\right)$.

In Week 2, consumer i's best guess of how much she will like movie j is a weighted average of the prior, her private signal, and the information that she obtains from her peers who have seen that movie, with weights that reflect the relative precision of these three pieces of information:

$$E_2[U_{ij} \mid X'_j\beta, s_{ij}, S_{ij2}] = \frac{h_j}{h_j + k_j + z_{i2}} X'_j\beta + \frac{k_j}{h_j + k_j + z_{i2}} s_{ij} + \frac{z_{i2}}{h_j + k_j + z_{i2}} S_{ij2}, \qquad (8)$$

where $z_{i2} = \frac{b_{i2}d}{b_{i2}+d}$.[5] The key implication is that in Week 2 the consumer has more information

4. The parameter σ_{j1} is $\sqrt{(1-\omega_j)^2(1/k_j + 1/d) + (1/r)}$.

5. The reason for having z_{i2} in this expression (and not b_{i2}) is that the equation predicts U_{ij}, not α^*_j. Therefore, we need to take into account not just the precision of S_{ij2} (b_{i2}) but also the precision of ν_{ij} (d).

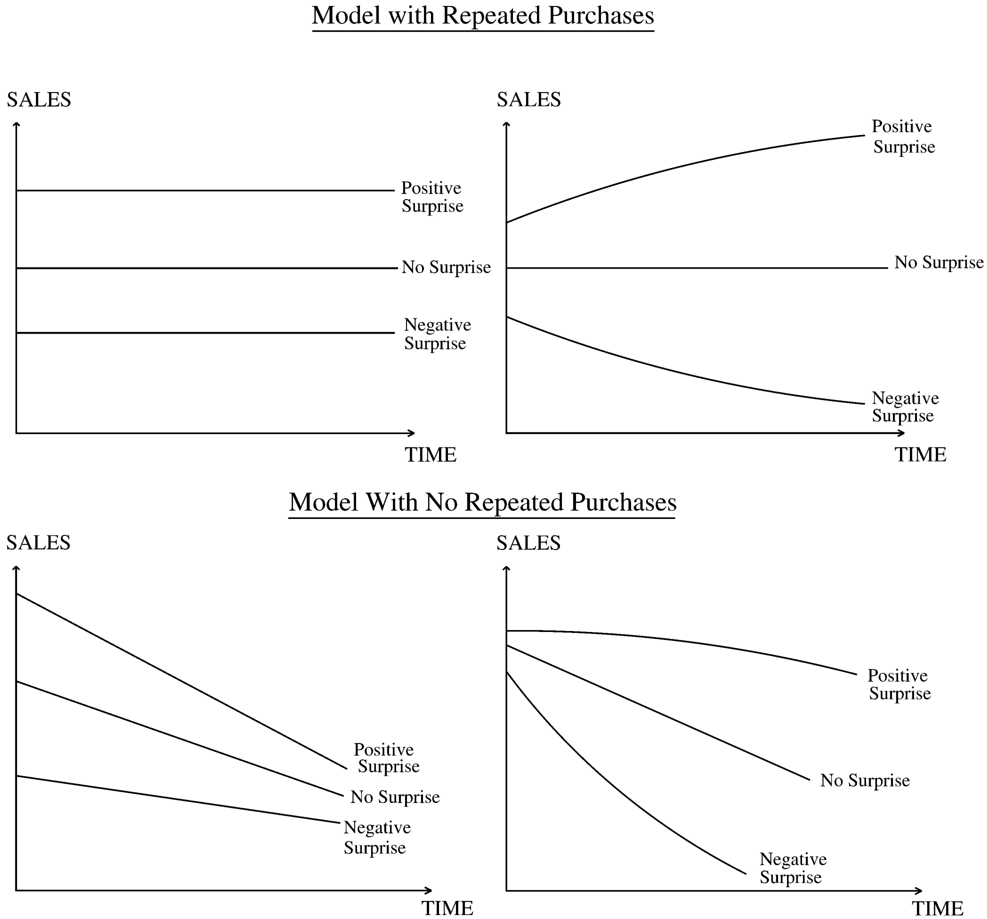


FIGURE 1

Theoretical changes in box-office sales over time. The upper panel shows the expected evolution of sales without social learning (left) and with social learning (right), in the model with no negative trend described in Section 2. The lower panel shows the expected evolution of sales in a model without social learning or with weak social learning (left) and with strong social learning (right), in the model with negative trends described in the Appendix 2

relative to the first week, and as a consequence the prior becomes relatively less important.[6] In each week after Week 2, more peer feedback becomes available. By iterating the normal learning model, in week t consumer i obtains an updated prediction of the utility provided by movie j:

$$
\begin{aligned}
E_t[U_{ij} \mid X'_j\beta, s_{ij}, S_{ij2}, S_{ij3}, \ldots, S_{ijt}] &= \omega_{j1t} X'_j\beta + \omega_{j2t} s_{ij} + \sum_{s=2}^{t} \omega_{j3s} S_{ijs} \\
&= \frac{h_j}{h_j + k_j + \sum_{s=2}^{t} z_{is}} X'_j\beta + \frac{k_j}{h_j + k_j + \sum_{s=2}^{t} z_{is}} s_{ij} + \sum_{s=2}^{t} \frac{z_{is}}{h_j + k_j + \sum_{s=2}^{t} z_{is}} S_{ijs}. \quad (9)
\end{aligned}
$$

6. Compare the weight on the prior in equation (5) with the weight on the prior in equation (8). It is clear that $\frac{h_j}{h_j+k_j} > \frac{h_j}{h_j+k_j+z_i}$.

The probability that the representative consumer sees the movie in week t is therefore[7]

$$P_t = \Phi\left(\frac{(1-w_{j1t})\alpha_j^* + w_{j1t}X_j'\beta - q}{\sigma_{jt}}\right). \tag{10}$$

Since individuals receive new feedback each week, the prior and the individual signal become less important over time and the true film quality becomes more important over time. It is clear from equation (10) that as t grows, sales are determined less and less by $X_j'\beta$ and more and more by α_j^* because w_{j1t} declines over time and $1-w_{j1t}$ increases over time.

I am interested in characterizing the change over time in the probability that a consumer decides to see a given movie as a function of the first-week surprise. For simplicity, consider first a movie whose *ex ante* expected value is equal to its average cost: $X'\beta = q$. Three cases are relevant.

(1) Positive surprise. If the true quality of the movie is higher than the prior ($\alpha_j^* > X_j'\beta$), then sales in the opening weekend are higher than the case of no surprise, and they grow over time. It is possible to show that

$$\frac{\partial P_t}{\partial t} > 0 \quad \text{if } \alpha_j^* > X_j'\beta. \tag{11}$$

This is shown in the bottom panel of Figure 1. The intercept shifts upward, reflecting the fact that the movie is a positive surprise in the opening weekend. The positive slope reflects the fact that consumers in Week 2 observe the utility of their peers in Week 1 and infer that the movie is likely to be better than their original expectation.

Furthermore, while we know from equation (9) that each passing week adds more information, the marginal value of this information declines over time. For example, the increase in information that a consumer experiences between Week 1 and Week 2 is larger than the increase in information that a consumer experiences between Week 2 and Week 3, given the information available in Week 2. And the latter is larger than the increase in information that a consumer experiences between Week 3 and Week 4, given the information available in Week 3.[8] It is possible to show that

$$\frac{\partial^2 P_t}{\partial t^2} < 0 \quad \text{if } \alpha_j^* > X_j'\beta. \tag{12}$$

This implies that the function that describes the evolution of ticket sales over time is concave.

(2) Negative surprise. The opposite is true if the true quality of the movie is lower than the prior ($\alpha_j^* < X_j'\beta$). In this case, the probability of seeing the movie is lower at the opening weekend and declining over time:

$$\frac{\partial P_t}{\partial t} < 0 \quad \text{if } \alpha_j^* < X_j'\beta. \tag{13}$$

This is shown in the bottom panel of Figure 1. Moreover, because the value of marginal information declines over time, the function that describes ticket sales is convex in time:

$$\frac{\partial^2 P_t}{\partial t^2} > 0 \quad \text{if } \alpha_j^* < X_j'\beta. \tag{14}$$

7. The term σ_{jt} is $\sqrt{w_{j2t}^2(1/d+1/k_j)+\left(\sum_{s=2}^t z_{is}\right)/\left(k_j+h_j+\sum_{s=2}^t z_{is}\right)^2+1/r}$.

8. When t is small, the consumer has limited information on α_j^* and the additional information provided by peers may be significant. But when t is large, the consumer has already obtained many estimates of α_j^* so that the marginal information provided by additional feedback is limited. Formally, the weight on the prior not only declines but also declines at a slowing rate: $\frac{\partial^2 w_{j1t}}{\partial t^2} < 0$.

(3) No surprise. If the true quality of the movie is equal to the prior ($\alpha_j^* = X'_j\beta$), then the probability of going to see the movie is on average constant over time: $P_t = P_1$. In this case, there is no surprise regarding the quality of the movie and social learning has no effect on ticket sales.

In the more general case where $X'_j\beta$ is not equal to q, the expression for the surprise is more complicated, but the intuition is similar: for a given prior $X'_j\beta$, a large enough α_j^* is associated with an increase in intercept and a positive slope, while a small enough α_j^* is associated with a decrease in intercept and a negative slope. Additionally, the inequalities in equations (12) and (14) also generalize.

The above model is kept purposefully simple. First, the model above assumes that the utility of watching a movie does not decline in the number of times a consumer watches it. As a result, it fails to capture an important feature of the dynamics of movie ticket sales, namely that sales of movies typically decline over time. Second, I have taken the timing of purchase as exogenous.[9] Both assumptions greatly simplify the model. In Appendix 2, I discuss how the predictions of the model might change if these two assumptions are relaxed.

One might wonder why, in my model, sales do not follow a logistic curve pattern, as it is often the case in models of technology diffusion. A logistic curve pattern would imply that the time profile of ticket sales is convex early on and concave later on. In other words, initially sales grow at an accelerating rate; once the product has reached a large market penetration, sales keep growing, but at a decreasing rate. Two opposite forces generate the logistic pattern. On one hand, the probability that consumer i buys the product increases if those near i have already bought it. This implies that increases in the number of those who have bought the product result in increases in the probability of buying by those who have not bought yet. On the other hand, the pool of potential new buyers shrinks over time as more and more individuals have already bought. This implies that the number of individuals at risk of buying declines over time. While the first force dominates early on (so that we see increases in the number of buyers at an accelerating rate), the second force dominates later on (so that we see increases in the number of buyers at a decelerating rate). A key feature of this model is therefore the assumption that a purchase by consumer i depends on purchases by other buyers near i. By contrast, my model does not have this feature. For simplicity, in my model, individuals update their expectation of the value of the movie based on the decisions of all their peers (see the definition of the peer signal, S_{ijt}, in Appendix 1). The shape of sale trends over time is not driven by the diffusion of information from nearby peers to peers further away. Instead, it is driven by the fact that the marginal increase in information on movie quality declines over time. Each passing week adds more information—as the number of peers who have seen the movie increases—but the marginal value of this information declines over time.

3. EMPIRICAL PREDICTIONS

The model above has several testable predictions that I bring to the data in Sections 6 and 7.

1. In the presence of strong enough social learning, sales of movies with stronger than expected opening weekend demand and sales of movies with weaker than expected opening weekend demand should *diverge* over time. This prediction is shown in Figure 1 and follows from equations (11) and (13). In the absence of social learning, or with weak social learning, we should see no divergence over time or even convergence.

9. In each period, individuals decide whether to see a particular movie by comparing its expected utility to the opportunity cost of time, q_{it}, assumed to be completely idiosyncratic. This assumption is not completely unrealistic because it says that individuals have commitments in their lives (such as work or family commitments) that are not systematically correlated with the opening of movies. On the other hand, it ignores the possibility that consumers might want to wait for uncertainty to be resolved before making a decision.

2. In the presence of social learning, the effect of a surprise should be stronger for movies with a more diffuse prior and weaker for movies with a more precise prior. Intuitively, when a consumer has a precise idea of whether she is going to like a specific movie (strong prior), the additional information provided by her peers should matter less relative to the case when a consumer has only a vague idea of how much she is going to like a movie (weak prior). Formally, this is evident from equation (9). A more precise prior (a larger h_j) implies a larger ω_{j1t} and therefore a smaller ω_{j3t} (everything else constant). This means that with a more precise prior, the additional information provided by the peers, S_{ijs}, will receive less weight, while the prior, $X'_j\beta$, will receive more weight. In the absence of social learning, there is no particular reason for why the correlation between sales trend and first-week surprise should vary systematically with precision of the prior.
3. In the presence of social learning, the effect of a surprise should be stronger for consumers who have larger social networks. The idea is that receiving feedback from 20 peers is more informative than receiving feedback from 2 peers. Formally, this is clear from equation (A.2) in Appendix 1. Equation (A.2) shows that a larger N_i implies a more precise estimate of movie quality based on peer feedback (*i.e.* a smaller variance of S_{ijt}). In the absence of social learning, there is no particular reason for why the correlation between sales trend and first-week surprise should vary systematically with size of the social network.
4. In the presence of social learning, the marginal effect of a surprise on sales should decline over time. For example, the amount of updating that takes place in Week 2 should be larger than the amount of updating that takes place in Week 3 given what is already known in Week 2. The implication is that the pattern of sales of movies with a positive (negative) surprise should be concave (convex) in time. This is evident from equations (12) and (14). In the absence of social learning, there is no particular reason for why the curvature of sales over time should vary systematically with the sign of first-week surprise.
5. In the presence of social learning, consumers should only respond to surprises that reflect new information on movie quality. They should not update their priors based on surprises that reflect factors other than movie quality. For example, consider the case of a movie whose opening weekend demand is weaker than expected because of bad weather. In this case, low demand in first week does not imply that the quality of the movie is low. Therefore, low demand in the first week should not lead consumers to update and should have no negative impact on subsequent sales.[10] In the absence of social learning, there is no particular reason for why variation in first-week demand due to surprises in movie quality and variation in first-week demand due to factors unrelated to quality should have different effects on sales trends.

4. DATA

I use data on box-office sales from the firm ACNielsen-EDI. The sample includes all movies that opened between 1982 and 2000 for which I have valid sales and screens data.[11] Besides total box-office sales by movie and week, it reports production costs, detailed genre classification, ratings, and distributor. I have a total of 4992 movies observed for 8 weeks. Total sample size is therefore $4992 \times 8 = 39,936$. This data set was previously used in Goettler and Leslie (2005).

10. Formally, one can think of weather shocks as part of the cost of going to see the movie, u_{it}. In the case of bad weather in the opening weekend, u_{i1} is high for many consumers.

11. I drop from the sample movies for which sales or number of screens are clearly misreported. In particular, I drop movies that report positive sales in a week but zero number of screens, or vice versa. I am interested in movies that are released nationally. Therefore, I drop movies that open only in New York and Los Angeles.

TABLE 1
Summary statistics

	(1)
Weekend sales (million)	1.78
	(4.37)
Weekend sales in opening weekend (million)	4.54
	(8.15)
Production costs (million)	16.88
	(0.81)
Number of screens	449.6
	(696.9)
Number of screens in opening weekend	675.7
	(825.5)
Favourable review	0.480
	(0.499)
Total TV advertising (million)	6.85
	(5.54)
Action	0.071
Adventure	0.018
Animated	0.027
Black comedy	0.017
Comedy	0.191
Documentary	0.047
Drama	0.340
Fantasy	0.011
Film festival	0.003
Horror	0.035
Musical	0.016
Romantic comedy	0.065
Sci-fiction	0.018
Sequel	0.068
Suspense	0.062
Western	0.005
Number of movies	4992

Notes: Standard errors are in parentheses. Dollar figures are in 2005 dollars. The sample includes 4992 movies that opened between 1982 and 2000. Each of these movies is observed for 8 weeks. Total sample size is 39,936. Sample size for reviews and TV advertising are 5064 and 14,840, respectively.

I augment box-office sales data with data on advertising and critic reviews. Unfortunately, these data are available only for a limited number of years. Data on TV advertising by movie and week were purchased from the firm TNS Media Intelligence. They include the totality of TV advertising expenditures for the years 1995–2000. Data on movie reviews were hand collected for selected years and newspapers by a research assistant. The exact date of the review and an indicator for whether the review is favourable or unfavourable were recorded. These data were collected for *The New York Times* for the movies opening in the years 1983, 1985, 1987, 1989, 1991, 1993, 1995, 1997, and 1999; for *The Wall Street Journal*, *USA Today*, *The Chicago Sun-Times*, *The Los Angeles Times*, *The Atlanta Journal-Constitution*, and *The Houston Chronicle* for the movies opening in the years 1989, 1997, and 1999; and for *The San Francisco Chronicle* for the movies opening in the years 1989, 1993, 1995, 1997, and 1999.

Summary statistics are in Table 1. The average movie has box-office sales equal to \$1·78 million in the average week. Box-office sales are higher in the opening weekend: \$3·15 million.

Production costs amount to \$4·54 million. All dollar figures are in 2005 dollars. The average movie is shown on 449 screens on average and on 675 screens in the opening weekend. The average movie has \$6·85 million in cumulative TV advertising expenditures. About half of the reviews are favourable. The bottom of the table shows the distribution of movies across genres. Comedy, drama, and action are the three most common genres.[12]

5. IDENTIFICATION OF SURPRISE IN OPENING WEEK DEMAND

The first step in testing the predictions of the model is to empirically identify surprise in opening weekend sales. I define the surprise as the difference between realized box-office sales and predicted box-office sales in the opening weekend, and I use the number of screens in the opening weekend as a sufficient statistic for predicted sales. Specifically, I use the residual from a regression of first-week log sales on log number of screens as my measure of movie-specific surprise. (In some specifications, I also control for genre, ratings, distribution, budget, and time of release.)

Number of screens is arguably a valid measure of the *ex ante* expectations of demand for a given movie because it is set by profit-maximizing agents (the theatre owners), who have a financial incentive to correctly predict consumer demand for a movie. Number of screens should therefore summarize the market expectation of how much a movie will sell based on all information available before opening day: cast, director, budget, advertising before the opening weekend, the quality of reviews before the opening weekend, the buzz in blogs, the strength of competitors, and any other demand shifter that is observed by the market before the opening weekend.

Deviations from this expectation can therefore be considered a surprise. These deviations reflect surprises in how much the concept of a movie and its cast resonate with the public. While theatres seem to correctly guess demand for movies *on average*, there are cases where the appeal of a movie and therefore its opening weekend demand are higher or lower than expected. These surprises are the ones used in this paper for identification. Formally, theatres seek to predict P_1 in equation (7). It is easy to show that in the case of a positive surprise in quality—*i.e.* when a movie true quality is higher than the prior ($\alpha_j^* > X'_j\beta$)—theatres' prediction, $\hat{P}_1$, is lower than realized sales: $\hat{P}_1 < P_1$. The opposite is true in the case of a negative surprise—*i.e.* when a movie true quality is lower than the prior ($\alpha_j^* < X'_j\beta$). In this latter case, $\hat{P}_1 > P_1$.[13]

Column 1 in Table 2 shows that the unconditional regression of log sales in first weekend on log screens in first weekend yields a coefficient equal to 0·89 (0·004), with R^2 of 0·907. This regression is estimated on a sample that includes one observation per movie ($N = 4992$). The

12. Not all movies have positive sales for the entire 8-week period. Because the dependent variable in the econometric models will be in logs, this potentially creates a selection problem. To make sure that my estimates are not driven by the potentially non-random selection of poorly performing movies out of the sample, throughout the paper I report estimates where the dependent variable is the log of sales +\$1. The main advantage of this specification is that it uses a balanced panel: all movies have non-missing values for each of the 8 weeks. I have also re-estimated my models using the selected sample that has positive sales and found generally similar results.

13. To see why, assume that theatres have the same information as consumers and use this information to predict P_1 in equation (7). The terms ω_j, $X'_j\beta$, q, and σ_{j1} are known, but α_j^* is unknown. Assume that theatres use the normal learning model to predict α_j^*: $E_1[\alpha_j^* \mid X'_j\beta, s_{ij}] = w_j X'_j\beta + (1 - w_j)s_{ij}$, where $w_j = m_j/(a_j + m_j)$ and $a_j = (dk_j)/(d + k_j)$. The weight on the prior used by theatres (w_j) is different from the weight on the prior used by consumers (ω_j in equation (5)). In particular, it is easy to see that $w_j > \omega_j$. This implies that even if consumers and theatres have the same information, theatres put more weight on the prior and less on their private signal. Intuitively, this is because theatres seek to predict α_j^*, while consumers seek to predict their own utility U_{ij}.

TABLE 2
Regression of first-weekend sales on number of screens

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log screens	0.892	0.895	0.882	0.870	0.802	0.806	0.813
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
R^2	0.907	0.908	0.909	0.912	932	0.936	0.937
Genre		Y	Y	Y	Y	Y	Y
Ratings			Y	Y	Y	Y	Y
Production cost				Y	Y	Y	Y
Distributor					Y	Y	Y
Weekday, month, week						Y	Y
Year							Y

Notes: Standard errors are in parentheses. The dependent variable is the log of opening weekend box-office sales. There are 16 dummies for genre, 8 dummies for ratings, 273 dummies for distributor, 18 dummies for year, 6 dummies for weekday, 11 dummies for month, 51 dummies for week. Sample size is 4992.

R^2 indicates that about 90% of the variation in first-week sales is predicted by theatre owners. Thus, about 10% of the variation cannot be predicted by theatre owners.[14]

Columns 2–7 show what happens to the predictive power of the model as I include an increasingly rich set of covariates. If my assumption is correct and number of screens is indeed a good summary measure of all the information that the market has available on the likely success of a movie, the inclusion of additional covariates should have limited impact on R^2. In Column 2, the inclusion of 16 dummies for genre has virtually no impact as R^2 increases from 0·907 to 0·908. Similarly, the inclusion of production costs and eight dummies for ratings raises R^2 only marginally, from 0·908 to 0·912. Including 273 dummies for the identity of the distributor, 12 dummies for months, 56 dummies for week, 6 dummies for weekday, and 18 dummies for year raises R^2 to 0·937. Overall, it is safe to say that the addition of all available controls has a limited impact on the fit of the model once the number of screens is controlled for.

My identification strategy is based on the assumption that theatre chains set the number of screens devoted to each movies based on their expectation of how well this movie will resonate with the public. As shown in my model, this assumption is consistent with profit maximization on the part of theatre owners. This assumption is also consistent with institutional features of this

14. There is ample anecdotal evidence from the movie industry that there is considerable uncertainty about first-week sales. Even the day before the opening, studio executives and theatre chain executives are typically unsure of how well a movie will perform. Friend (2009) reports that “Opening Fridays are always tense: at Fox, they used to call the long hours ‘the Vigil’, and Universal’s Adam Fogelson [an executive at Universal] says that openings are ‘an exciting, nauseating thrill ride’". These anecdotal accounts are consistent with my assumption that demand is uncertain, even just days before the opening. Additionally, there are countless newspaper articles and websites devoted to predictions of opening weekend box-office sales and at least two websites that allow betting on opening weekend box-office sales. (Unfortunately, betting sites did not exist during the period for which I have sales data.) During production, studios do use focus groups to determine which aspects of the story are more likely to resonate with the public and how to best tailor advertising to different demographic groups. The unpredictable component of demand that I focus on here is very different in nature as it takes place just before release. I am not aware of focus group analysis performed after production and marketing are completed to predict first-weekend sales.

industry.[15] Moreover, the assumption is not inconsistent with a Mincer–Zarnowitz test. Specifically, I test whether the estimated prediction error is systematically correlated with features of the movie that are known to theatres at or before the opening weekend. If my assumption is correct, the residual from the regressions shown in Table 2 should be orthogonal to movie characteristics (and function of movie characteristics) that are known to theatres when they set the number of screens. I consider a fifth-order polynomial in cumulative advertising expenditures measured at the opening weekend, a fifth-order polynomial in the fraction of positive critic reviews among those that have been released at or before the opening weekend, indicators for movie genre, indicators for month of the release, and indicators for day of the week of the release.[16] A regression of prediction error on these variables and predicted sales indicates that these variables are individually not statistically different from zero at conventional levels. This would suggest that the difference between predicted demand and actual demand is not systematically correlated with variables that are known when the prediction is made. However, the test for joint significance is less clear-cut. The test statistics for the joint test is 1·32, and the p-value is 0·101, right on the margin. To address this potential concern, in the empirical analysis below I show that my empirical findings are generally quite robust to alternative definition of surprise. In particular, all the models used to estimate surprise in Table 2 yield similar empirical estimates of social learning, indicating that the addition of controls over and above the number of screens does not affect my findings.

One might still wonder whether it is possible that theatres might find it profitable to set the number of screens not equal to the expected demand. For example, would it be optimal to systematically lower the number of screens to artificially boost surprise in first-week demand? It seems unlikely. First, consumers would probably discount systematic underscreening and would adjust their expectations accordingly. More fundamentally, number of screens is simply a summary measure that I use to quantify expectations on demand. In my model, consumers' expected utility is based on movie underlying characteristics (director, actors, genre, etc.) as well as their private signal. For a given set of movie characteristics and for a given signal, manipulation of the number of screens would have no impact on consumer demand. The reason is that consumers in Week 2 do not respond *causally* to surprises in first-week demand. They respond causally only to variation in movie quality. Manipulating surprises in first-week demand without changing actual movie quality is unlikely to generate higher demand and higher profits.

Table 3 shows the distribution of surprises in the opening weekend box-office sales, together with some examples of movies. For example, the entry for 75% indicates that opening weekend sales for the movie at the 75th percentile are 46% higher than expected. The distribution appears symmetric, and it is centred around 0·02. Since surprise is a regression residual, its mean is zero by construction.

An example of a movie characterized by large positive surprise is "The Silence of the Lambs". Before the opening weekend, it was expected to perform well, since it opened on 1479 screens, substantially above the average movie. But in the opening weekend, it significantly exceeded expectations, totalling sales of about $25 million. In this case, sales were significantly higher than the amount theatres were expecting based on the screens assigned to it. Other examples

15. I have contacted the 10 largest theatre chains in America and asked them who sets the number of screens and how. Four of these chains responded to my enquiry. Based on their responses, it seems that the decision about the number of screens is usually made by a special department within the movie chain, often called the "Film Department". Notably, the movies to be screened and the number of screenings are set based on how well the film department expects the movie will perform. For example, one of the chains (Rave Motion Picture) explicitly said that the decision about screening is made by one man named Eric Bond in their film department, and it is "based on his expectations of how good a movie will be".

16. Of course, the equation for surprise used for this test does not include any of these variables.

TABLE 3
Distribution of surprises in opening weekend sales

Percentile (1)	Surprise (2)	Examples (3)
5%	−1.21	Tarzan and the Lost City; Big Bully; The Fourth War
10%	−0.87	Tom & Jerry ; The Phantom of the Opera; Born to Be Wild
25%	−0.41	Home Alone 3; Pinocchio; Miracle on 34th Street
Median	0.02	Highlander 3; The Bonfire of the Vanities; Fear and Loathing in Las Vegas
75%	0.46	Alive; Autumn in New York; House Party
90%	0.84	The Muse; Sister Act; Forrest Gump; Tarzan
95%	1.08	Breakin; Ghostbusters; The Silence of the Lambs

Notes: Entries represent the distribution of surprises (in percent terms) in opening weekend box-office sales. Mean surprise is 0 by construction.

of movies that experienced significant positive surprises are “Ghostbusters”, “Sister Act”, and “Breakin”. More typical positive surprises are represented by movies in the 75th percentile of the surprise distribution, such as “Alive”, “Who Framed Roger Rabbit”, and “House Party”. For many movies, the demand in the first week is close to market expectations. Examples of movies close to the median include “Highlander 3”, “The Bonfire of the Vanities”, “The Sting 2”, and “A Midsummer Night’s Dream”. Examples of negative surprise include “Home Alone 3”, “Pinocchio”, “Lassie”, and “The Phantom of the Opera”. These four movies opened on a large number of screens (between 1500 and 1900) but had first weekend box-office sales lower than one would expect based on the number of screens. Interestingly, there are two very different versions of Tarzan movies. One is an example of a strong negative surprise (“Tarzan and the Lost City”), while the second one is a strong positive surprise (“Tarzan”).

6. EMPIRICAL EVIDENCE

I now present empirical tests of the five implications of the model described in Section 3. I begin in Subsection 6.1 with tests of Prediction 1. In Subsection 6.2, 6.3, and 6.4, I present tests of Predictions 2, 3, and 4, respectively. Later, in Section 7, I discuss the interpretation of the evidence and I present a test of Prediction 5.

6.1. *Prediction* 1*: surprises and sale dynamics*

6.1.1. Baseline estimates. The main implication of the model is that in the presence of social learning, movies with a positive surprise in first weekend sales should have a slower rate of decline in sales than movies with a negative surprise in first weekend sales (Prediction 1). In the absence of social learning, movies with a positive and negative surprise should have the same rate of decline in sales.

Figure 2 shows a graphical test of this prediction based on the raw data. It shows unconditional average log sales by week and surprise status. The upper line represents the decline in average sales for movies with a positive surprise, and the bottom line represents the decline for movies with a negative surprise. The pattern shown in the figure is striking. Consistent with Prediction 1, movies with a positive surprise experience a slower decline in sales than movies with a negative surprise. As a consequence, the distance between the average sales of positive and negative surprise movies is relatively small in the opening weekend but increases over time. After 8 weeks, the difference is much larger than in Week 1.

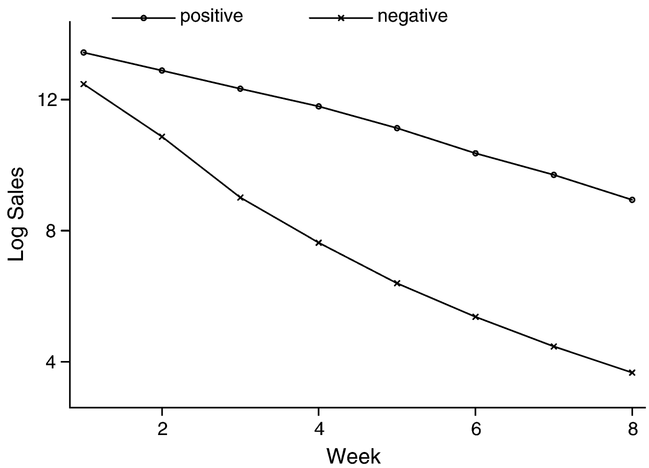


FIGURE 2

The decline in box-office sales over time by opening weekend surprise. The figure plots average log box-office sales by week for movies that experienced a positive first weekend surprise and for movies that experienced a negative first weekend surprise. Sales are in 2005 dollars. The sample includes 4992 movies

To test whether the difference in slopes between positive and negative surprise movies documented in Figure 2 is statistically significant, I estimate models of the form

$$\ln y_{jt} = \beta_0 + \beta_1 t + \beta_2 (t \times S_j) + d_j + u_{jt}, \tag{15}$$

where $\ln y_{jt}$ is the log of box-office sales in week t; S_j is surprise or an indicator for positive surprise; and d_j is a movie fixed effect. Identification comes from the comparison of the *change* over time in sales for movies with a positive and a negative surprise. To account for the possible serial correlation of the residual within a movie, standard errors are clustered by movie throughout the paper.

The coefficient of interest is β_2. A finding of $\beta_2 > 0$ is consistent with the social learning hypothesis since it indicates that the rate of decline of sales of movies with positive surprise is slower than the rate of decline of sales of movies with negative surprise, as in Figure 2. A finding of $\beta_2 = 0$, on the other hand, is inconsistent with the social learning hypothesis since it indicates that the rate of decline of sales of movies with positive and negative surprise is the same.

It is important to note that the interpretation of β_2 is not causal. The model in Section 2 clarifies that, under social learning, a stronger than expected demand in Week 1 does not *cause* a slower decline in sales in the following weeks. A stronger than expected demand in Week 1 simply indicates (to the econometrician) that the underlying quality of the movie is better than people had expected. It is the fact that movie quality is better than expected and the diffusion of information about movie quality that *cause* the slower decline in sales in the weeks following a positive surprise in first-week demand. A positive surprise in first-week demand is simply a marker for better than expected quality.

Estimates of variants of equation (15) are in Table 4. In Column 1, I present the coefficient from a regression that only includes a time trend. This quantifies the rate of decay of sales for the average movie. The entry indicates that the coefficient on t is $-0{\cdot}926$. In Column 2, the regression includes the time trend $\times$ surprise, with surprise defined as the residual from a regression of log sales on number of screens, indicators for genre, ratings, production cost, distribution, week, month, year, and weekday. (This definition of surprise is the one used in

TABLE 4
Decline in box-office sales by opening weekend surprise

	(1)	(2)	(3)	(4)
t	−0.926	−0.926	−1.258	
	(0.012)	(0.011)	(0.017)	
t × surprise		0.463		
		(0.016)		
t × positive surprise			0.616	
			(0.022)	
t × bottom surprise				−1.320
				(0.019)
t × middle surprise				−0.984
				(0.020)
t × top surprise				−0.474
				(0.017)
R^2	0.77	0.79	0.79	0.79

Notes: Standard errors are clustered by movie and displayed in parentheses. Each column reports estimates of variants of equation (15). The dependent variable is log weekly box-office sales. All models include movie fixed effects. Surprise refers to deviation from predicted first-weekend sales. By construction, surprise has mean 0. In Column 4, "bottom surprise" is an indicator for whether the movie belongs to the bottom tercile of the surprise distribution (most negative surprise). Similarly, "middle surprise" and " top surprise" are indicators for whether the movie belongs to the middle or top (most positive) tercile of the surprise distribution, respectively. Sample size is 39,936.

Column 7 of Table 2.) The coefficient β_2 from this regression is equal to 0·46 and is statistically different from zero. Since the variable "surprise" has by construction mean zero, the coefficient on t is the same in Columns 1 and 2.

In Column 3, S_j is an indicator for whether surprise is positive. The entry for β_2 quantifies the difference in rate of decay between positive and negative surprise movies shown graphically in Figure 2. The entry indicates that the rate of decline for movies with a negative surprise is $-1{\cdot}25$, while the rate of decline for movies with a positive surprise is about half as big: $-1{\cdot}25 + 0{\cdot}619 = -0{\cdot}63$. This difference between positive and negative surprise movies is both statistically and economically significant.

In Column 4, I divide the sample into three equally sized groups depending on the magnitude of the surprise, and I allow for the rate of decline to vary across terciles. Compared to the model in Column 3, this specification is less restrictive because it allows the rate of decline to vary across three groups, instead of two. I find that the rate of decline is a monotonic function of surprise across these three groups. The coefficient for the first tercile (most negative surprise) is $-1{\cdot}32$. The coefficient for the second tercile (zero or small surprise) is $-0{\cdot}98$. The coefficient for the third tercile (most positive surprise) is $-0{\cdot}47$.[17]

To better characterize the variation in the rate of decline of sales, I estimate a more general model that allows for a movie-specific decline:

$$\ln y_{jt} = \beta_0 + \beta_{1j} t + d_j + u_{jt}, \tag{16}$$

17. To see whether the advent of Internet has changed the amount of social learning, I have re-estimated the models allowing for a different slope in the years after 1995. The results do not indicate that the amount of social learning was larger in the years 1995–2000 relative to the years 1982–1994 possibly because online networks were still limited.

where the rate of decline β_{1j} is now allowed to vary across movies. When I use estimates of equation (16) to compare the distribution of β_{1j} for positive and negative surprise movies, I find that the distribution of the slope coefficients for movies with a positive surprise is more to the right than the distribution of the slope coefficients for movies with a negative surprise. The 25th percentile, the median, and the 75th percentile are, respectively, $-0{\cdot}96$, $-0{\cdot}41$, and $-0{\cdot}11$ for positive surprise movies, and $-1{\cdot}91$, $-1{\cdot}23$, and $-0{\cdot}60$ for negative surprise movies.

6.1.2. Advertising. One important concern is that the difference in sales trends between positive and negative surprise movies may be caused by changes in advertising expenditures induced by the surprise. If studios initially set their optimal advertising budget based on the expected performance of a movie, then it is plausible that a surprise in actual performance will change their first-order conditions. In particular, if studios adjust their advertising expenditures based on first-week surprise, estimates in Table 4 may be biased, although the sign of the bias is *a priori* undetermined.[18]

In practice, directly controlling for TV advertising does not appear to significantly affect estimates.[19] This is shown in Table 5. In the top part of the table, I report estimates of models where I interact time with the first-week surprise (as in Column 2 of Table 4), while in the bottom part I report estimates of models where I interact time with an indicator for positive surprise (as in Column 3 of Table 4). Column 1 reproduces estimates of the baseline specification that does not include controls for advertising from Table 4. Column 2 shows estimates of the same baseline specification obtained using the subsample of movies for which I have advertising data. The comparison of estimates in Columns 1 and 2 suggests that the subsample of movies for which I have advertising data generates estimates that are qualitatively similar to the full sample estimates.

In Column 3, I include controls for TV advertising broadcasted in the current week and the previous 4 weeks, allowing the effect of advertising to be different depending on how far back in time it takes place. Specifically, in week t, I separately control for the logarithm of total expenditures for TV advertising in weeks t, $t-1$, $t-2$, $t-3$, and $t-4$ or earlier. This specification allows the elasticity of sales to advertising in a given week to be different depending on when the advertising takes place relative to the current week. As one may expect, more recent advertising has a bigger impact on sales than older advertising. The coefficients and standard errors on log advertising are 0·037 (0·005), 0·019 (0·004), 0·018 (0·005), 0·017 (0·005), and 0·0005 (0·005) for weeks t, $t-1$, $t-2$, $t-3$, and $t-4$, respectively. These coefficients indicate that advertising that is older than 3 weeks has a limited impact on current sales. The coefficients are shown graphically in the top panel of Figure 3.

It is in principle possible that four lags may not be enough to fully control for the effect of advertising. In Column 4, I include controls for TV advertising broadcasted in the current week and the previous 10 weeks. Specifically, in week t, I separately control for the logarithm of total expenditures for TV advertising in weeks t, $t-1$, $t-2, \ldots, t-9$, and $t-10$ or earlier. The coefficients on current and lagged advertising are shown graphically in the bottom panel of Figure 3. As for the case of four lags, more recent advertising has a bigger impact on sales than older advertising, and coefficients for lag further than $t-3$ are not statistically significant.

The models in Columns 3 and 4 assume that the elasticity of sales to advertising is the same for all types of movies. In Columns 5–9, I relax this assumption by allowing advertising in a

18. The sign of the bias depends on whether the marginal advertising dollar raises revenues more for positive or negative surprise movies.

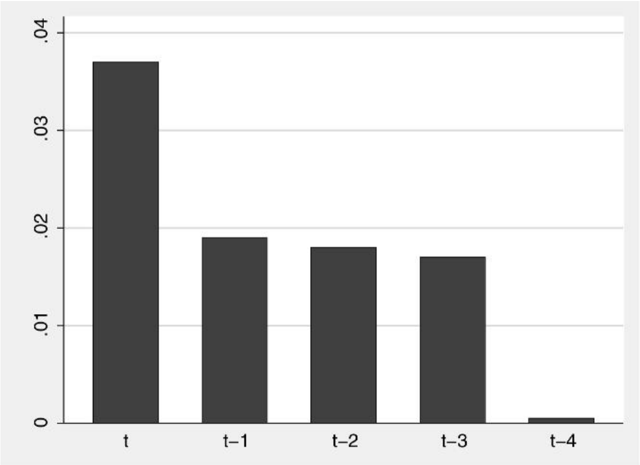
19. I do not have data for newspaper advertising. However, TV advertising typically represent at least 80% of advertising expenditures for movies.

TABLE 5
The effect of controlling for TV advertisement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Model 1										
t	−0.926	−0.903	−0.769	−0.745	−0.779	−0.770	−0.753	−0.763	−0.751	−0.835
	(0.011)	(0.028)	(0.036)	(0.067)	(0.071)	(0.071)	(0.070)	(0.072)	(0.103)	(0.124)
t × surprise	0.463	0.667	0.607	0.609	0.618	0.625	0.640	0.638	0.638	0.654
	(0.016)	(0.040)	(0.040)	(0.040)	(0.041)	(0.042)	(0.043)	(0.043)	(0.046)	(0.052)
R^2	0.79	0.81	0.82	0.83	0.84	0.84	0.85	0.85	0.85	0.86
Model 2										
t	−1.258	−1.260	−1.059	−1.092	−1.122	−1.124	−1.114	−1.123	−1.189	−1.264
	(0.017)	(0.055)	(0.059)	(0.082)	(0.084)	(0.084)	(0.084)	(0.086)	(0.113)	(0.138)
t × positive surprise	0.616	0.740	0.672	0.661	0.677	0.677	0.689	0.676	0.653	0.654
	(0.022)	(0.062)	(0.059)	(0.060)	(0.063)	(0.063)	(0.062)	(0.063)	(0.066)	(0.082)
R^2	0.79	0.80	0.82	0.82	0.83	0.83	0.83	0.84	0.84	0.85
Full sample	Y									
TV Ads (current week + 4 lags)			Y							
TV Ads (current week + 10 lags)				Y	Y	Y	Y	Y	Y	
TV Ads (current week + 10 lags) × genre dummies					Y	Y	Y	Y	Y	Y
TV Ads (current week + 10 lags) × big budget						Y	Y	Y	Y	Y
TV Ads (current week + 10 lags) × ratings dummies							Y	Y	Y	Y
TV Ads (current week + 10 lags) × month dummies								Y	Y	Y
TV Ads (current week + 10 lags) × after opening									Y	Y
Network TV Ads (current week + 10 lags)										Y
Cable TV Ads (current week + 10 lags)										Y
Local TV Ads (current week + 10 lags)										Y
Syndicated TV Ads (current week + 10 lags)										Y

Notes: Standard errors are clustered by movie and displayed in parentheses. All models include movie fixed effects. Column 1 reproduces the baseline estimates from Table 4. Controls in Column 3 include the logarithm of total expenditures for advertising in weeks t, $t-1$, $t-2$, $t-3$, and $t-4$ or earlier. Controls in Column 4 include the logarithm of total expenditures for advertising in weeks t, $t-1$, $t-2$, $\ldots$, $t-9$, and $t-10$ or earlier. Controls in Columns 5–9 include the interactions of advertising in each lag with the relevant observable. Sample size is 39,936 in Column 1 and 5975 in Columns 2–10.

given week to affect sales differently depending on the interaction of type of movie and the amount of time since advertising was broadcasted. I begin in Column 5 by allowing heterogeneity in the effect of advertising across genres. Specifically, I control for the logarithm of total



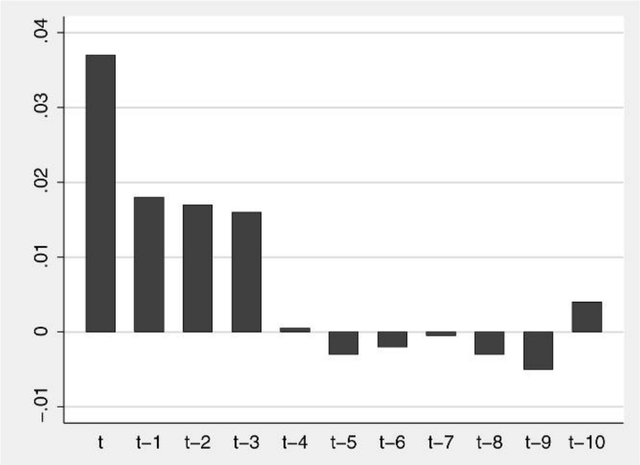


FIGURE 3

The effect of advertising in a given week on sales in week t. The top panel shows the coefficients on the logarithm of total expenditures for TV advertising in weeks t, $t-1$, $t-2$, $t-3$, and $t-4$ or earlier in a regression where the dependent variable is sales in week t and other covariates include a time trend and a time trend interacted with surprise (specification in Column 3 in Table 5, top panel). Only the coefficients on t, $t-1$, $t-2$, and $t-3$ are statistically different from zero. The bottom panel shows the coefficients on the logarithm of total expenditures for TV advertising in weeks t, $t-1$, $t-2$, ..., $t-9$, and $t-10$ or earlier in a regression where the dependent variable is sales in week t and other covariates include a time trend and a time trend interacted with surprise (specification in Column 4 in Table 5, top panel). Only the coefficients on t, $t-1$, $t-2$, and $t-3$ are statistically different from zero

expenditures for TV advertising in weeks t, $t-1$, $t-2$, ..., $t-9$, and $t-10$ or earlier and for the interactions of each of these lags with 16 indicators for genre. This specification not only allows for the elasticity of sales in a given week to advertising to be different depending when the advertising takes place but also for each lag, it allows the elasticity of sales to advertising of, say, comedies, to differ from the elasticity of sales to advertising for, say, dramas.

In Column 6, in addition to all the controls of Column 5, I also control for the interactions of log advertising in weeks t, $t-1$, $t-2$, ..., $t-9$, and $t-10$ or earlier with an indicator for whether the movie has an above-average production budget. This specification allows for the elasticity of sales in a given week to advertising to be different depending on whether the movie is a big or a small production. In Column 7, I allow for heterogeneity in the effect of advertising across movies with different ratings. Specifically, I add the interactions of log advertising in weeks $t, t-1, t-2, \ldots, t-9$, and $t-10$ or earlier with eight indicators for ratings. In Column 8, I allow for seasonal variation in the effect of advertising. Specifically, I add the interactions of log advertising in weeks t, $t-1$, $t-2$, $\ldots$, $t-9$, and $t-10$ or earlier with 12 indicators for month. In Column 9, I allow for the effect of advertising in a given week on sales in a later week to vary depending on whether advertising takes place before or after the opening week. In particular, I add the interactions of log advertising in weeks $t, t-1, t-2, \ldots, t-9$, and $t-10$ or earlier with an indicator for before and after the opening week.

In Columns 1–9, my measure of advertising is the total expenditures for TV advertising. My advertising data separately identify expenditures for cable TV, network TV, local TV, syndicated TV, and other TV expenditures. In Column 10, I control separately for different types of TV advertising. Specifically, I add among the controls log advertising in weeks t, $t-1$, $t-2$, $\ldots$, $t-9$, and $t-10$ for each of these categories.

The comparison of the coefficients in Columns 3–10 with the baseline estimate in Column 2 indicates that a rich set of controls for TV advertising and a very flexible functional form that includes up to 10 lags and interactions with several observables have limited impact on the estimates. I have also run specifications where I control for cumulative advertising. Specifically, cumulative expenditures for advertising for movie j in week t represent the sum of expenditures for all TV ads broadcasted for movie j until week t. Estimates (not reported in the table) do not change significantly.[20]

Overall, the table suggests that changes in advertising do not appear to be a major factor in explaining my results. Why does controlling for advertising does not change my estimates significantly? Estimates of β_2 in equation (15) that do not control for advertising are upward biased only to the extent that the advertising that occurs in the weeks *after the opening weekend* is positively correlated with surprise. However, most advertising takes place *before* the release of a movie. One reason for this is that the typical distribution contract in this industry ensures that the studios—who pay for most of the advertising—receive a higher share of profits from earlier week sales than later week sales. Given the depreciation over time in the advertising effect documented above, studios have limited incentives to advertise in later weeks. Indeed, in my sample, 94% of TV advertising occurs before the opening day. Furthermore, the endogenous response of advertising to surprise, while positive, is quantitatively small.[21] In other words, (i) only a small amount of advertising is at risk of being affected by the surprise and (ii) the elasticity of advertising to first-week surprise is empirically small. Because of these two features, the endogenous reaction of advertising to surprise does not appear to represent an important source of bias in practice.

I also note that, even if advertising could explain the slower decline in sales for positive surprise movies, it does not explain the comparative statics results on the precision of the prior and the size of the social network that I describe in Subsections 6.2 and 6.3 below.

6.1.3. Critic reviews. Another potentially important omitted variable is represented by critic reviews.[22] The concern is that movie critics react to a surprise in opening weekend by

20. Specifically, the coefficients on time and time $\times$ surprise are $-0{\cdot}996$ (0·024) and 0·661 (0·036), respectively. The coefficients on time and time $\times$ an indicator for positive surprise are $-1{\cdot}329$ (0·048) and 0·730 (0·053), respectively.

21. A regression of log advertising in Weeks 2–8 on first-week surprise yields a coefficient that is statistically significant but small in magnitude. The coefficient is 0·18 (0·02).

22. Reinstein and Snyder (2005) estimate the effect of movie reviews on attendance.

TABLE 6
The effect of controlling for reviews

	Baseline full sample		Baseline partial sample		Controlling for reviews partial sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
t	−0.926	−1.258	−0.891	−1.232	−0.943	−1.279	−1.060	−1.391
	(0.011)	(0.017)	(0.018)	(0.029)	(0.018)	(0.029)	(0.028)	(0.035)
t × surprise	0.463		0.543		0.534		0.499	
	(0.016)		(0.026)		(0.025)		(0.027)	
t × positive surprise		0.616		0.660		0.649		0.605
		(0.022)		(0.037)		(0.036)		(0.038)
Favourable reviews					5.114	5.184	3.959	3.847
					(0.408)	(0.437)	(0.449)	(0.475)
Favourable reviews × t							0.289	0.332
							(0.047)	(0.047)
R^2	0.79	0.79	0.79	0.78	0.80	0.79	0.80	0.80
N	39,936	39,936	14,840	14,840	14,840	14,840	14,840	14,840

Notes: Standard errors are clustered by movie and displayed in parentheses. The dependent variable is log weekly box-office sales. Columns 1 and 2 report baseline estimates (reproduced from Columns 2 and 3 of Table 4). Columns 3 and 4 report estimates of the same models obtained using the sample for which data on movie reviews are available. Columns 5–8 include controls for reviews and are based on the same sample used for Columns 3 and 4. "Favourable reviews" is the fraction of positive reviews among all the reviews published until the relevant week. All models include movie fixed effects.

covering unexpected successes. This could have the effect of boosting sales for positive surprise movies, thus generating the difference in rate of decline between positive and negative surprise movies documented above. Like advertising, the majority of reviews take place *before* the release of a movie. In my data, 85% of newspaper reviews are published at or before the date of the opening.

Directly controlling for positive reviews does not affect estimates significantly. This is shown in Table 6. As for advertising, data on reviews are available only for a subset of movies. Columns 1 and 2 report baseline estimates reproduced from Table 4. In Columns 3 and 4, I report estimates of the baseline model obtained using the subsample for which I have data on reviews. In Columns 5 and 6, I control for the cumulative share of reviews that are favourable as a fraction of all the reviews published until the relevant week.

The comparison of Columns 5 and 6 with Columns 3 and 4 suggests that the inclusion of control reviews has limited impact on my estimates. The coefficients on time and time × surprise change from −0·856 and 0·509 (Column 3) to −0·927 and 0·494 (Column 5), respectively. The coefficients on time and time × an indicator for surprise change from −1·163 and 0·603 (Column 4) to −1·227 and 0·582 (Column 6), respectively. The coefficient on favourable reviews is positive and significantly different from zero, although it cannot necessarily be interpreted causally. I have also re-estimated the models in Columns 5 and 6 separately for each newspaper and found similar results.

Finally, in Columns 7 and 8 I control for the cumulative share of reviews that are favourable and for the interaction of this cumulative share with t. This specification allows for the effect of good reviews to vary over time. The coefficient on the interaction of review and t indicates that positive reviews slow down the decline in sales. The coefficients of interest are not affected significantly.

Like for advertising, I also note that if the only reason for a slowdown in the rate of decline of positive surprise movies were critic reviews, we would not necessarily see the comparative statics results on the precision of the prior and size of the social network that I describe in Subsections 6.2 and 6.3 below.

6.1.4. Supply effects and sold-out movies. So far, I have implicitly assumed that all the variation in sales reflects consumer demand. However, it is possible that changes in the availability of screens affect sales. Consider the case where there is no social learning, but some theatre owners react to the first-week surprise by adjusting the movies they screen. This type of supply effect has the potential to affect sales, especially in small towns, where the number of screens is limited. For example, in Week 2, a theatre owner in a small town may decide to start screening a movie that had a positive surprise elsewhere, thereby increasing the number of customers who have access to that movie.

This is important because it implies that the evidence in Table 4 may be explained not by learning on the part of consumers but by learning on the part of theatre owners. To test for this possibility, I have re-estimated my models using sales *per screen* as the dependent variable. In this specification, the focus is on changes in the average number of viewers for a given number of screens. These results are therefore not affected by changes in the number of theatres screening a given movie. I find that estimates of the effect of a surprise are qualitatively robust to the change in the definition of the dependent variable.[23] Note that the interpretation of this specification requires caution. Number of screens is an endogenous variable, which presumably adjusts as a function of demand shocks. If there is social learning, a positive surprise in Week 1 will result in an increase in demand in the following weeks, and, as a consequence, it will also cause an increase in the number of screens devoted to that particular movie. While this specification is helpful in testing for the possibility of supply effects, it is not my preferred specification because, by focusing on sales per screen, it discards useful variation in the dependent variable.

Another possible interpretation of Table 4 is that the slower decline in sales for positive surprise movies reflects sold-out movies, not social learning. Suppose, *e.g.* that demand in Week 1 exceeds capacity and that some of this excess demand gets shifted to later weeks. To test this possibility directly, I re-estimate equation (15) excluding movies that might have sold out. In particular, I re-estimate my models dropping movies that are in the top 1%, 3%, or 5% of the per-screen attendance distribution. Estimates indicate that results are robust to this specification. (See Table 8 in Moretti, 2008). Furthermore, this alternative explanation is not consistent with the comparative statics on precision of the prior and of the signal documented in Subsections 6.2 and 6.3 below. If the only thing driving results is that some positive surprise movies are sold out, we should not expect to see the effect of surprise vary systematically with the precision of the prior.

6.1.5. Robustness. Here I probe the robustness of estimates in Table 4 using two sets of alternative specifications. First, I investigate whether my estimates are sensitive to changes in the definition of surprise. In Table 4, surprise was the residual in a regression of log sales on number of screens and controls, including 16 dummies for genre, 8 dummies for ratings, cost of production, and controls for timing of the opening (18 dummies for year, 12 dummies for month, 52 dummies for week of the year, and 7 dummies for day of the week). If the number of screens is a good measure of the market expectations of the demand for a movie, the presence

23. For example, when the dependent variable is sales, the rate of decline of a positive surprise movie is half of the rate of decline of a negative surprise movie. When the dependent variable is sales per screen, the rate of decline of a positive surprise movie is 65% of the rate of decline of a negative surprise movie. See Table 5 in Moretti (2008) for details

TABLE 7
Robustness of the estimated effect of surprise to the inclusion of controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel 1: additional controls are included in surprise equation							
Model 1							
$t\times$ surprise	0.422	0.424	0.420	0.416	0.445	0.442	0.438
	(0.012)	(0.012)	(0.012)	(0.012)	(0.015)	(0.016)	(0.017)
Model 2							
$t\times$ positive surprise	0.672	0.666	0.659	0.646	0.585	0.587	0.591
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.024)
Surprise equation controls for							
Genre		Y	Y	Y	Y	Y	Y
Ratings			Y	Y	Y	Y	Y
Production cost				Y	Y	Y	Y
Distributor					Y	Y	Y
Special weekends						Y	Y
Week							Y
Panel 2: additional controls are included in sales equation							
Model 3							
$t\times$ surprise	0.422	0.422	0.425	0.427	0.444	0.441	0.437
	(0.012)	(0.012)	(0.012)	(0.012)	(0.015)	(0.015)	(0.015)
Model 4							
$t\times$ positive surprise	0.672	0.671	0.673	0.674	0.684	0.679	0.671
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.024)	(0.024)
Sales equation controls for							
$t\times$ genre		Y	Y	Y	Y	Y	Y
$t\times$ ratings			Y	Y	Y	Y	Y
$t\times$ production cost				Y	Y	Y	Y
$t\times$ distributor					Y	Y	Y
$t\times$ special weekends						Y	Y
$t\times$ week							Y

Notes: In panel 1, each entry is an estimate of the parameter β_2 in equation (15), where the definition of surprises varies across columns. Surprise refers to deviation from predicted first-week sales, where predicted first-week sales are obtained using the number of screens as a predictor and an increasing number of controls, as specified at the bottom of the panel. Panel 2 reports estimates of equation (17), where an increasing number of controls are added to the sales equation, as specified at the bottom of the panel. In this panel, surprise is defined using only the number of screens as predictors. All models include movie fixed effects. Sample size is 39,936.

of these additional covariates should have no effect on estimates. Panel 1 of Table 7 shows that this is indeed the case.

I report estimates of β_2 when surprise is defined as the residual from a regression of first weekend log sales on the log of number of screens in the opening weekend and a varying set of controls. The table shows that alternative definitions of surprise yield very similar estimates of the coefficient β_2. For example, in Column 1 surprise is defined as the residual in a regression of sales on number of screens only. The point estimates are 0·422 in Model 1 and 0·672 in Model 2. These estimates differ only slightly from the baseline estimates in Table 4, which are 0·463 and 0·616. Adding dummies for genre, ratings, production costs, and distributor has limited impact (Columns 2–5).

As a second check, I investigate whether estimates are sensitive to the addition of controls in the sale equation (equation (15)), holding fixed the definition of surprise. All time-invariant

TABLE 8
Precision of the prior and precision of peers' signal

	Precision of prior		Precision of signal	
	Sequel (1)	Variance of surprise (2)	Movies for teenagers (3)	Size of the release (4)
t	−1.259 (0.017)	−1.368 (0.313)	−1.163 (0.022)	−1.180 (0.019)
t × positive surprise	0.627 (0.023)	−0.789 (0.431)	0.576 (0.029)	0.568 (0.028)
t × sequel	0.015 (0.075)			
t × positive surprise × sequel	−0.190 (0.096)			
t × variance		0.156 (0.445)		
t × positive surprise × variance		2.003 (0.612)		
t × teen movie			−0.224 (0.034)	
t × positive surprise × teen movie			0.091 (0.046)	
t × screens				−0.133 (0.024)
t × positive surprise × screens				0.094 (0.028)
R^2	0.79	0.79	0.79	0.79

Notes: Standard errors are clustered by movie and displayed in parentheses. The dependent variable is log weekly box-office sales. Columns 1 and 2 report estimates of variants of equation (18). In Column 1, precision of the prior is measured by sequel status. Movies that are sequels are expected to have more precise priors. In Column 2, precision of the prior is measured by the variance of the first weekend surprise in box-office sales. Genres with a larger variance are expected to have less precise priors. Column 3 reports an estimate of equation (19), where "Teen Movie" is an indicator equal to one if the intended audience is teenagers, based on detailed genre. Teenagers are expected to have a larger social network and therefore more social learning. Column 4 reports an estimate of equation (20), where "Screens" is the number of screens in the opening weekend. Movies with a larger number of screens are expected to send a more precise signal. Number of screens is divided by 1000 to make the reported coefficients easier to read. All models include movie fixed effects. Sample size is 39,936.

movie characteristics are fully absorbed in this equation by movie fixed effects, but one might still be concerned that the rate of decline differs across movies and is correlated with surprise. For example, one might be concerned that the rate of decline for, say, adventure movies is slow, and at the same time adventure movies tend to have positive surprises. I do not expect this to be a likely scenario. If screens are indeed a good measure of market expectations, they should account for all systematic differences across genres and other movie characteristics.

To address this concern, I investigate whether estimates are sensitive to the inclusion of an increasing number of film characteristics interacted with time trends

$$\ln y_{jt} = \beta_0 + \beta_1 t + \beta_2 (t \times S_j) + \beta_3 (t \times X_j) + d_j + u_{jt}, \tag{17}$$

where X_j includes genre, ratings, production costs, distributor, and date of the release (year, month, day of the week), and surprise is based only on the number of screens. In this model, any systematic differences in the rate of decline in sales across different genres, ratings, costs, distributors, and time of release are accounted for.

Panel 2 of Table 7 indicates that estimates are robust to the inclusion of these additional controls. For example, the model in Column 2 controls for a genre-specific rate of decline in sales. Identification of β_2 arises because surprises vary within genre. The coefficients, 0·422 and 0·671, are remarkably similar to the baseline coefficients. The models in Columns 3–5 allow the rate of decline to vary depending on ratings, budget, and distributor.

6.1.6. Endogenous competitors. Movies rarely open alone. Often there are two or three movies opening in the same weekend. One concern is that there are certain holiday weekends that are characterized by particularly high demand and other weekends that are characterized by low demand so that quality and number of movies released in a given weekend vary endogenously. Specifically, Einav (2007) has shown that multiple movies of good quality are released in high-demand weekends and fewer movies with low average quality are released in low-demand weekends. In other words, the number of competitors is not random, and the quality of competitors is also not random. One may be concerned that in these cases the mismatch between screens and realized demand does not reflect only consumers' surprise.

Two pieces of evidence suggest that in practice this may not be a significant source of bias. First, I explicitly identify 16 holiday weekends and other special weekends when demand is particularly high, including Presidents Day, Labour Day, Memorial Day, July 4th, Thanksgiving, Christmas, New Years, Valentines Day, Martin Luther King Day, Columbus Day, Veterans Day, Halloween, Cinco de Mayo, Mothers Day, and Superbowl Sunday. In Column 6 of Table 7, I present models that include separate controls for each of these special weekends and the interaction between the special weekend indicators with the time trend. In these models, the rate of change over time of sales of movies released on, say, Labour Day weekend is allowed to be different from the rate of change over time of sales of movies released on, say, the 4th of July and is also allowed to be different from the rate of change over time of sales of movies released on an average weekend. Systematic differences between movies released in high- and low-demand weeks should be controlled for. I add these controls either as covariates in the sales equations (upper panel) or as controls in the surprise equation (lower panel).

Second, in Column 7, I present models that include controls for 52 indicators for week of the year and the interaction between the 52 indicators for week of the year with the time trend. These models are even more general than the models that account for holiday weekends. In these models, the rate of change over time of sales of a movie released, say, in Week 5 is allowed to be different from the rate of change over time of sales of a movie released, say, in Week 6. Therefore, systematic differences between movies released in different parts of the year are controlled for.

More in general, I note that there are reasons to expect that the mere presence of competitors is not necessarily a problem for my identification strategy. My identification is based on surprises. Since the identity of competitors is not a surprise, the number of screens allocated to a given movie should incorporate the best prediction of demand given its competitors. The variation in sales caused by the presence of competitors should not be reflected in my measure of surprise. A movie facing a competitor that is expected to be strong will have a smaller number of screens than a movie facing a competitor expected to be weak. Take the 4th of July weekend, *e.g.*. The fact that there are many high-quality movies on the 4th of July weekend does not imply that movies that open on the 4th of July weekend will have a systematically low or high surprise. While competitors are not random, the deviation from expectations is arguably random. Indeed, the empirical evidence is consistent with this notion. When I test for whether surprises are systematically different on holiday weekends, I do not find that my measure of surprise is systematically larger or smaller on holiday weekends or special weekends. Moreover, the fact that controlling for holiday weekends and special weekends does not affect my estimates confirms that the endogenous sorting of high-quality movies into these weekends is not the driving force behind my estimates.[24]

6.2. *Prediction* 2*: precision of the prior*

The evidence uncovered so far seems consistent with one of the predictions of the social learning hypothesis. But if such evidence is indeed explained by social learning, there are several additional implications that one should see in the data. In the rest of this section and in the next one, I test four additional implications of the social learning hypothesis.

Prediction 2 indicates that social learning should matter more for movies with a diffuse prior than for movies with a precise prior. To test this prediction, I estimate models of the form

$$\ln y_{jt} = \beta_0 + \beta_1 t + \beta_2 (t \times S_j) + \beta_3 (t \times \text{precision}_j) + \beta_4 (t \times S_j \times \text{precision}_j) + d_j + u_{jt}, \quad (18)$$

where precision_j is a measure of the precision of the prior for movie j. The coefficient of interest is the coefficient on the triple interaction between the time trend, the surprise, and the precision of the prior, β_4, which is predicted to be negative in the presence of social learning. It is important to note that, unlike equation (15), this model does not compare movies with positive and negative surprise. Instead, for a given surprise, this model tests whether the effect of the surprise is systematically associated with the precision of the prior.

To empirically identify which movies have precise priors and which have diffuse priors, I propose two measures. First, I use a dummy for sequels. It is reasonable to expect that consumers have more precise priors for sequels than non-sequels. For example, after having seen the movie "Rocky", consumers are likely to have a more precise idea of whether they will like its sequel "Rocky II". Second, to generalize this idea, I calculate the variance of the first-week surprise in box-office sales by genre. Genres with large variance in first-week surprise are characterized by more quality uncertainty and therefore consumers should have more diffuse priors on their quality. Indeed, sequels are the genre with the second smallest variance in first-week surprise.[25]

Table 8 indicates that the data are consistent with the prediction of the model. In Column 1, the coefficient on the triple interaction is negative and statistically significant. Consider two movies, identical except for the fact that the first is a sequel and the second is not. Suppose that they both have the same positive surprise in first-week sales. The estimates in Column 1 imply that the surprise has more impact on the rate of decline of the second than the first presumably because the additional information provided by the surprise matters less for the sequel. Quantitatively, the rate of decline of sales of the second movie implied by Column 1 is $-0{\cdot}63$, while the rate of decline of the first movie is $-0{\cdot}81$. In other words, consistent with the social learning hypothesis, the same positive surprise benefits sales of the second movie significantly more.

In Column 2, I use the variance in first-week surprise as a measure of precision of the prior. In this case, I expect the coefficient on the triple interaction to be positive since higher variance means a less precise prior. Indeed, this is the case empirically.

24. One might still be concerned that a positive surprise competitor might mechanically induce a negative surprise for a given movie. While this is likely, this mechanism alone would not generate the increased divergence in sales over time between positive surprise movies and negative surprise movies documented in Section 6. It would simply result in a parallel shift in sales, without necessarily affecting the rate of decline. Moreover, this mechanism alone would not generate the comparative statics in Sections 6.2 and 6.3.

25. The genre with the smallest variance is Western. The genre with the largest variance is Action.

6.3. *Prediction* 3*: size of the social network*

Prediction 3 indicates that social learning should be stronger for consumers who have a larger social network since these consumers receive more precise feedback from their peers than consumers with small social networks. While I do not have direct information on the size of the social network, it seems plausible to assume that teenagers have a more developed social network than older adults. Social learning should therefore be more important for teen movies. To test this prediction, I estimate models of the form

$$\ln y_{jt} = \beta_0 + \beta_1 t + \beta_2 (t \times S_j) + \beta_3 (t \times \text{teen}_j) + \beta_4 (t \times S_j \times \text{teen}_j) + d_j + u_{jt}, \qquad (19)$$

where teen_j is an indicator variable for whether the movie's target audience is teenagers. Here the coefficient of interest is the coefficient on the triple interaction, β_4, which is predicted to be positive in the presence of social learning. Like equation (18), this model does not compare movies with positive and negative surprises. Instead, for a given surprise, this model tests whether the effect of the surprise is larger for teen movies.

Estimates in Column 3 of Table 8 are consistent with Prediction 3. The coefficient on the triple interaction is indeed positive, indicating that a positive surprise has a larger impact on rate of decline of sales for teen movies than non-teen movies.

A related approach uses the size of the movie release. The precision of the feedbacks that consumers receive from friends should be larger for movies opening in many theatres than for movies opening in few theatres. For example, a positive surprise for a movie that opens in 2000 theatres should generate a more precise signal and therefore more updating than an identically sized surprise for a movie that opens only in 20 theatres. To test this hypothesis, in Column 4 I estimate the following model:

$$\ln y_{jt} = \beta_0 + \beta_1 t + \beta_2 (t \times S_j) + \beta_3 (t \times \text{screens}_j) + \beta_4 (t \times S_j \times \text{screens}_j) + d_j + u_{jt}, \qquad (20)$$

where screens_j is the number of screens in which the movie opened. The expectation is that β_4 is positive. Consistent with the hypothesis, the coefficient on the triple interaction in Column 5 of Table 8 is positive.

6.4. *Prediction* 4*: does learning decline over time?*

Prediction 4 in Section 3 involves the time path of the diffusion of information. It indicates that in the presence of social learning a positive surprise movie should generate a concave sales profile and a negative surprise movie should generate a convex sales profile. To test this prediction, I estimate the following model:

$$\ln y_{jt} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 (t \times 1(S_j > 0)) + \beta_4 (t^2 \times 1(S_j > 0)) + d_j + u_{jt}, \qquad (21)$$

where $1\ (S_j > 0)$ is a dummy for positive surprise movies. If it is empirically the case that the sale profile is concave for positive surprise movies, I should find that the second derivative is negative: $2(\beta_2 + \beta_4) < 0$. Similarly, if the profile is convex for negative surprise movies, I should find that the second derivative is positive: $2\beta_2 > 0$.

My estimates are indeed consistent with this prediction. Point estimates and standard errors are as follows: β_1 $-1{\cdot}88$ (0·051); β_2 0·089 (0·006); β_3 1·38 (0·060); β_4 $-0{\cdot}110$ (0·007). R^2 is 0·79. Statistical tests confirm that the curvature for positive surprise movies is concave and the curvature for negative surprise movies is convex. A test of the hypothesis that $2(\beta_2 + \beta_4) < 0$ has p-value $= 0{\cdot}0001$. A test of the hypothesis that $2\beta_2 > 0$ also has p-value $= 0{\cdot}0001$.

7. NETWORK EXTERNALITIES

Overall, the evidence in the previous section is consistent with the hypothesis that the diffusion of information through social learning significantly affects consumers' purchasing decisions. The rate of decline of sales for movies with positive surprise is much slower than the rate of decline of movies with negative surprise. This finding does not appear to be driven only by omitted variables. For example, this finding is not due to endogenous changes in advertising expenditures or critic reviews in the weeks after the opening. Similarly, this finding cannot be explained by endogenous changes in the number of screens devoted to a movie because it is robust to using per-screen sales as the dependent variable instead of sales.

Moreover, if this finding were only due to omitted variables, we would not necessarily observe the comparative static results based on precision of the prior and size of the social network. The effect of a surprise appears more pronounced when consumers have diffuse priors on movie quality and less pronounced when consumers have strong priors, as in the case of sequels. The effect of a surprise is also larger for movies that target audiences with larger social networks. Additionally, the amount of learning appears to decline over time. This fact is also consistent with the social learning hypothesis.[26]

In this section, I discuss an alternative explanation of the evidence, namely the possibility of network externalities. A network externality arises when a consumer's utility of watching a movie depends on the number of peers who have seen the movie or plan to see the movie. The social learning model assumes that individuals care about others' actions only because they convey information about the quality of a product. However, in addition to this informational externality, there may be direct pay-off interactions in the form of positive consumption externalities. While the social learning model assumes that individuals care about others' actions only because they convey information about the quality of a product, network externalities imply that each consumer's utility from a good depends *directly* on the consumption by others. This type of network effect may occur if, *e.g.* consumers draw utility from talking about a movie with their friends and that utility is increasing in the number of friends who have seen the movie. This possibility could be particularly relevant for certain subgroups—like teenagers—for whom social pressure, fashions, and conformity are important.

Network externalities can in principle generate all the four pieces of evidence described in Section 6.[27] This alternative interpretation is difficult to distinguish from social learning, and I cannot completely rule it out. However, I provide a simple test based on Prediction 5 in Section 3 that sheds some light on the relevance of this interpretation. Specifically, I test whether consumers respond to surprises in first-week sales that are orthogonal to movie quality,

26. It is in theory possible that the patterns in Table 4 are explained by a slow adjustment of consumers to the innovation represented by the surprise. Consider, *e.g.* a positive surprise movie, and assume that there is a fraction of consumers who cannot immediately go when a new movie is released but have already decided to see the film in later weeks. In this case, the positive surprise movie would experience a slower rate of decline in sales even in the absence of social learning. However, while this explanation could in theory generate the patterns documented in Table 4, it is not consistent with the comparative statics on precision of the prior and the signal documented below in Subsections 6.2 and 6.3. In the absence of social learning, there is no reason why we should see that the effect of the surprise is systematically associated with the precision of the prior or the size of the social network.

27. Consider, *e.g.* the evidence in Table 4. A movie with a positive surprise in Week 1 may attract more viewers in Week 2 not because of learning but because marginal consumers in Week 2 realize that a larger than expected number of their friends have seen the movie, making the movie more valuable to them. Moreover, if utility of watching a movie is a function of the *expected* number of friends who will ever see the movie, the network effects story is consistent not only with Table 4 but also with the comparative statics in Subsection 6.2. A consumer who uses the normal learning model to update her predictions on how many peers will ever see the movie should be more affected by a surprise in Week 1 when she has a diffuse prior relative to the case when she has a precise prior.

like weather shocks. Under social learning, lower than expected demand in Week 1 due to bad weather should have no significant impact on sales in the following weeks because demand shocks driven by weather do not reflect movie quality. On the other hand, under the network effects hypothesis, lower than expected demand in Week 1 caused by bad weather should still lower sales in the following weeks. If a consumer draws utility from talking about a movie with friends, she cares about how many friends have seen a movie, irrespective of their reasons for choosing to see it.

To implement this test, I estimate equation (15) by 2SLS, where I instrument S_j using weather in the opening weekend. By using weather as an instrument, I isolate the variation in surprise that only comes from weather shocks. This variation is arguably independent of the quality of a movie. The coefficient of interest is the 2SLS estimate of β_2. The network hypothesis predicts that $\beta_2 > 0$. By contrast, the social learning hypothesis predicts that $\beta_2 = 0$.

A major limitation of my data is that it only includes information on nationwide sales. However, it is possible to predict first-week surprise at the national level using weather conditions in seven large cities: New York, Boston, Chicago, Denver, Atlanta, Kansans City, and Detroit.[28] In particular, I measure variation in weather using maximum temperature, minimum temperature, precipitation, and snowfall, by city and day. Each of the seven cities has at least one weather monitor. For cities where more than one monitor is available, I average across all available monitors. I assign weather to a movie based on the weather in the day of the release. In the first stage, I regress (nationwide) sales on weather in each of the seven cities on the day of the release and the day before the release. In general, rain, snow, and colder temperatures in these seven cities are associated with significantly lower first-week demand and significantly lower first-week surprise.

Results are shown in Table 9. The first row corresponds to a specification where S_j is the surprise of movie j, while the second row corresponds to a specification where S_j is a dummy for whether the surprise of movie j is positive. Column 1 reports the baseline OLS estimates for the sample for which I have non-missing temperature data for all seven cities. These estimates are very similar to the corresponding estimates for the full sample. Column 2 reports instrumental variables estimates where instruments include minimum and maximum temperature on the opening day and the day before the opening day. It is possible that the effect of temperature is non-linear. For example, when it is cold, we expect higher temperatures to be associated with higher sales, but when it is hot, we expect higher temperatures to lower sales. For this reason, in Column 3 instruments also include the squares of minimum and maximum temperature on the opening day and the day before. In Column 4, instruments include minimum and maximum temperature as well as precipitation and snowfall on the opening day and the day before. In Column 5, instruments include minimum and maximum temperature, precipitation, and snowfall, on the opening day and the day before, as well as these variables squared. Surprise is defined as the deviation of sales from expected sales based on the number of screens. The last row in Table 9 corresponds to a test of whether the weather variables are jointly significant in the first stage. While only some of the first-stage coefficients are individually significant, taken together they are statistically significant. Tests for whether the first-stage coefficients are jointly equal to 0 have p-values below 0·0001 in all four columns, although the F test statistics are low.

Comparison of Columns 1 and 2 indicates that the coefficient on the surprise $\times\ t$ interaction drops from 0·413 to −0·107, while the coefficient on the positive surprise dummy $\times\ t$ interaction drops from 0·643 to −0·277. The corresponding coefficients in Column 2 are 0·085 and 0·123 and are not statistically significant. Columns 3 and 4 display similar results. Not surprisingly, the

28. Dahl and DellaVigna (2009) are the first to document that weather in selected cities shifts aggregate movie sales.

TABLE 9
Test of the network hypothesis

	OLS	2SLS			
		IV is minimum temperature, maximum temperature		IV is minimum temperature, maximum temperature, precipitation, snowfall	
	(1)	(2)	(3)	(4)	(5)
t × surprise	0.413	−0.107	0.085	−0.039	0.118
	(0.014)	(0.122)	(0.083)	(0.093)	(0.067)
t × positive surprise	0.643	−0.277	0.123	−0.129	0.181
	(0.025)	(0.235)	(0.152)	(0.184)	(0.125)
F test: first-stage coefficient = 0		3.54	3.07	2.82	2.60
p-Value		0.000	0.000	0.000	0.000
N	31,528	31,528	31,528	30,320	30,320
Weather enters linearly		Y		Y	
Quadratic in weather			Y		Y

Notes: Standard errors are clustered by movie and displayed in parentheses. Each entry is a separate regression and represents an estimate of the parameter β_2 in equation (15). The dependent variable is log weekly box-office sales. Column 1 reports OLS estimates based on the sample for which I have data on maximum and minimum temperature. Column 2 reports instrumental variables estimates where instruments include minimum and maximum temperature on the opening day and the day before the opening day in seven metropolitan areas. In Column 3, instruments include minimum and maximum temperature on the opening day and the day before the opening day as well as these variables squared. In Column 4, instruments include minimum and maximum temperature, precipitation, and snowfall on the opening day and the day before the opening day. In Column 5, instruments include minimum and maximum temperature, precipitation, and snowfall on the opening day and the day before the opening day as well as these variables squared. Surprise is defined based on the number of screens. Sample size varies because data on weather are missing for some cities in some days.

instrumental variable (IV) estimates are considerably less precise than the OLS estimates most likely because I cannot perform the analysis at the city level. For this reason, it is difficult to draw firm conclusions. However, the small, insignificant, and often wrong-signed point estimates in Columns 2–5 suggest that network effects cannot fully explain the large effects documented in Table 4.

8. ECONOMIC IMPORTANCE OF SOCIAL LEARNING

In this section, I use my estimates to quantify the economic magnitude of the social learning effect in the movie industry. To quantify the cumulative effect of social learning on sales, I compute how much higher (lower) are total sales of a movie that has a positive (negative) surprise when consumers learn from each other, relative to the counterfactual where the movie has the same quality but consumers do not learn from each other.

The cumulative effect on sales generated by social learning over the life of a movie is illustrated in Figure 4. The effect of social learning for a movie with positive surprise is represented by the shaded area, which is the area between the line representing the sale profile for a movie with a positive surprise and the counterfactual line. The counterfactual line has the same slope as a movie with no surprise. The effect of social learning should not include the fact that a positive surprise movie has a higher intercept because the higher intercept reflects the fact that a positive surprise movie has higher quality and therefore sells more *irrespective of social learning.* It is only the indirect effect that operates through the change in slope that reflects social learning. (For simplicity, in this calculation I ignore the non-linearity uncovered in Subsection 6.4.)

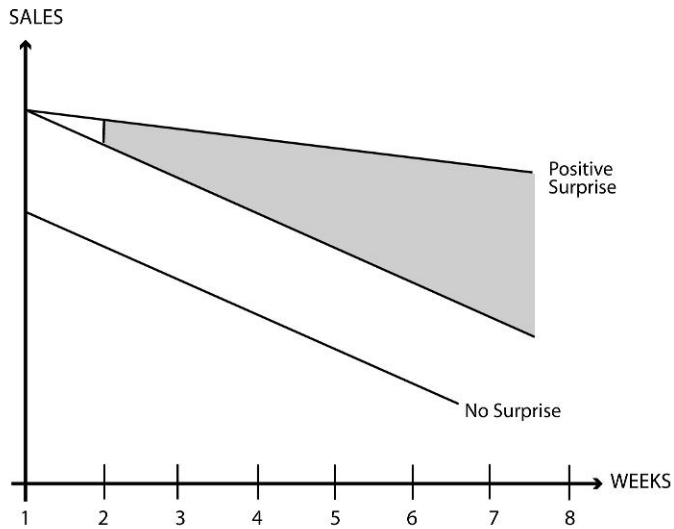


FIGURE 4

Effect of social learning on sales. The shaded area represents the increase in sales due to social learning for a movie with a positive surprise relative to a the average movie. The area is defined by the line representing the sale profile for a movie with a positive surprise and the counterfactual line. The counterfactual line has the same slope as a movie with no surprise and a higher intercept. The higher intercept reflects the fact that a positive surprise movie has higher quality and therefore sells more *irrespective of social learning.* The difference in slopes reflects the strength of social learning

I use estimates of the parameters in equation (15) to calculate the magnitude of the effect. The results of this exercise are striking. Consider first the typical movie with positive surprise—*i.e.* the one at the 75th percentile of the surprise distribution, with sales 46% above expectations. Social learning raises box-office sales by \$4·5 million over the lifetime of the movie. This effect is remarkably large, especially when compared with total sales. In particular, the social learning effect amounts to about 32% of total sales—relative to the case of a movie that has similar quality but where consumers do not communicate.[29] Since the distribution of surprises is slightly asymmetric, the effect of social learning on lifetime sales is slightly different for negative surprise movies. For the typical negative surprise movie—*i.e.* the movie at the 25th percentile of the surprise distribution, with sales 41% below expectations—social learning accounts for \$4·9 million in lost sales (or 34% of total sales).

By comparison, the effect of TV advertising on sales accounts for 48% of total sales. (This figure comes from a model that includes t, $t\times$ surprise, and TV advertising in weeks t, $t-1$, $t-2,\ldots,t-9$, and $t-10$ or earlier. The coefficients on current and lagged advertising are shown graphically at the bottom panel of Figure 3.) In other words, the effect of social learning for the typical movie appears to be about two-thirds as large as the overall effect of TV advertising. I view this as a remarkably large magnitude. This finding is important since it indicates that attracting a new consumer has a significant multiplier effect on sales because it increases the demand of other consumers.

A normative implication is that studios should advertise unexpected successes more often than they currently do. Empirically, advertising is not very sensitive to surprises. (The coefficient is positive, but the magnitude is small.) This is in part due to the fact that the typical distribution

29. The effect is slightly larger for movies that target audiences with large social networks, like teenagers.

contract in this industry ensures that the studios—who pay for most of the advertising—receive a high share of profits from earlier week sales and a low share of profits for later week sales. Thus, studios have limited incentives to advertise after the opening week. A change in the structure of these contracts may be beneficial for the industry because it could allow for more advertising of positive surprises and presumably higher sales.[30]

Moreover, these findings are relevant for marketing in this industry. Technological innovations promise to make "peer to peer" advertising increasingly important. For example, Facebook has recently unveiled plans to introduce advertising opportunities across the social networking site through word-of-mouth promotions.[31] Presumably, this type of innovative marketing can succeed only for products for which social learning and peer effects are important. It appears that movies are one example of such products.[32]

9. CONCLUSION

This paper makes two contributions. Substantively, this is among the first studies to credibly test for social learning using real-world, industry-wide data. There is a large and influential theoretical literature on the topic of social learning and informational cascades, but the empirical evidence is limited. Most of the existing empirical evidence is from a growing number of studies based on laboratory experiments. While laboratory experiments may be useful, real-world data are necessary to establish how important social learning and informational cascades are in practice.

I find that social learning is an important determinant of sales in the movie industry. Social learning makes successful movies more successful and unsuccessful movies more unsuccessful. Consistent with a simple model where consumers update their priors based on their peers' purchasing decisions, the rate of decline of movies with stronger than expected first-week demand is about half the rate of decline of movies with weaker than expected first-week demand.

While I cannot completely rule out alternative explanations,[33] the weight of the evidence is consistent with the social learning hypothesis. As the model predicts, the effect of a surprise on subsequent sales is smaller for movies for which consumers have strong priors and larger for movies for which consumers have more diffuse priors. Additionally, the effect of a surprise is more pronounced for groups of consumers who have more developed social networks. Finally, consumers do not seem to respond to surprises caused by factors that are orthogonal to movie quality, like weather shocks.

Quantitatively, social learning appears to have an important effect on profits in the movie industry. For the typical movie with positive surprise, social learning raises box-office sales by \$4·5 million—or about 32% of total revenues—relative to the case of a movie that has similar quality but where consumers do not communicate. The existence of this large "social multiplier" indicates that the elasticity of aggregate movie demand to movie quality is significantly larger than the elasticity of individual demand to quality.

30. This is not uncommon in other industries. For example, publishers often advertise the fact that a book is an unexpected best-seller.

31. Mark Zuckerberg, chief executive and founder of Facebook, says such peer recommendations are the "Holy Grail of advertising" and suggested that "nothing influences people more than a recommendation from a trusted friend".

32. Indeed, Yahoo is already experimenting with "peer to peer" advertising of movies using its user groups. In the long run, the introduction of this new type of marketing has the potential to further increase the amount of social learning and therefore make the difference in sale trends between positive surprise and negative surprise movies that I document even more pronounced.

33. The finding of the Mincer–Zarnowitz test are not definitive on whether the prediction error made by theatres is orthogonal to characteristics of the movies that are observed. While each of the observables taken individually appears to be not statistically different from zero, I cannot completely reject that, taken jointly, they are statistically significant.

The second contribution of the paper is methodological. This paper shows that it is possible to identify social interactions using aggregate data and intuitive comparative statics. In situations where individual-level data and exogenous variation in peer group attributes are not available, this approach is a credible alternative methodology for identifying spillovers and social interactions. There are many industries where the proposed methodology could be useful. Books, *e.g.* are experience goods whose quality is *ex ante* uncertain. Because of this uncertainty, it is conceivable that the probability that a consumer purchases a particular book is influenced by the purchasing decisions of others. Indeed, publishers often advertise the fact that a book is a best-seller.[34] One could use the number of copies in the first print as a measure of *ex ante* market expectations and test whether deviations in sales from the first print generate sales dynamics consistent with social learning. Additional tests would include testing for (i) differences in the precision of the prior (*e.g.* books by a first-time author should be characterized by more diffuse priors than books by an established author); (ii) differences in the size of the reader social network; and (iii) non-linearities in the functional form of sales trends.

Other consumption goods where social learning is likely to be important and where one could use my methodology are computer software, cars, designer clothes, and other fashion products.[35] Social learning is also important in the restaurant industry. Consumers who are unfamiliar with a restaurant often identify its quality by the fraction of seats occupied. Consumption of financial products is also likely to be influenced by social learning. The slogan for a recent campaign advertisement for Charles Schwab is “So far this year investors have opened 300,000 new Schwab accounts. Do they know something you don’t?”. Visa new advertising campaign has the slogan “More People Go With Visa”.

Social learning is not confined to consumers, but it is likely to be important for firms as well. Adoption of new technologies is a good example. When the return on an innovation is uncertain, late adopters may learn from earlier adopters, so that there is an informational externality associated with adopting early on.[36] One could use my methodology to compare the speed of adoption of innovations that have returns higher than expected for early adopter with the speed of adoption of innovation that have returns lower than expected. The comparative statics that I propose based on the precision of the prior, the size of the social network, and the concavity/convexity apply to this context.

Social learning may also be relevant for financial markets. For example, initial public offerings (IPO) of equity are often characterized by sharp increases in price in the days following the IPO. Why are IPOs on average severely underpriced by issuing firms? Welch (1992) posits that early adoptions induced by the low price may help start a positive cascade. My methodology could in principle be used to test this hypothesis. The test would involve showing that stronger than expected performance of a stock in the first day of trading increase its performance later. Additionally, the comparative statics tests that I propose here all apply to this context. For example, the precision of prior can be measured empirically by how well-known the CEO is or by the variance in the changes in first-day trading price by industry.

Social learning may also be relevant for voting. If candidate quality is difficult to observe in advance, and all voters receive a noisy signal on candidate quality, voters could use other people’s voting behaviour to update their quality expectations. Using a methodology similar to the one proposed here, Knight and Schiff (2007) find that social learning is particularly important in presidential primaries, where voters in different states vote sequentially.

34. It is common, *e.g.* to see book covers with the statement “100,000 copies sold”. Amazon even provides information on additional books that were considered by customers who bought a particular book.

35. For example, a recent advertisement campaign for Ipod simply states “100 Million Ipods Sold”.

36. Diffusion models similar to the one developed here were used to document the spreading of new technologies based on peer imitation (Griliches, 1957; Bass, 1969).

APPENDIX 1

Here I describe formally how a consumer uses her peer feedbacks to obtain an estimate of movie quality. Of the N_i of consumer i's peers, n_i see the movie in Week 1 and communicate to consumer i their *ex post* utility: U_{pj}, for $p = 1, 2, \ldots, n_i$, where p indexes peers. To extract information on quality from peers' feedback, consumer i needs to take into account the fact that the set of peers from whom she receives feedback is selected. These are the peers who *ex ante* found the movie appealing enough that they decided to see it: they tend to have a high signal s_{pj}. Consumer i receives a feedback from peer p only when $E_1[U_{pj} \mid X_j'\beta, s_{pj}] > q_{p1}$ (equation (6)). If she ignored this selection and simply averaged the feedbacks $U_{1j}, U_{2j}, \ldots, U_{nj}$, consumer i would obtain a biased estimate of the quality of the movie.[37]

In Week 2, consumer i obtains an estimate of quality, α_j^*, from the observed $U_{1j}, U_{2j}, \ldots, U_{n_i j}$ and the number of peers who have not seen the movie, by maximizing the following maximum likelihood function:

$$L_{ij2} = L\left[U_{1j}, U_{2j}, \ldots, U_{n_i j}, n_i \mid \alpha_j^*\right] = \prod_{p=1}^{n_i} \int_q^{\infty} f(U_{pj}(\alpha_j^*), V)dV \prod_{p=n_i+1}^{N_i} \Pr\{V_{pj} < q\} \tag{A.1}$$

$$= \prod_{p=1}^{n_i} \sqrt{d}\phi(\sqrt{d}(U_{pj} - \alpha_j^*))\left(1 - \Phi\left(\frac{q - \omega_j X_j'\beta - (1-\omega_j)U_{pj}}{\sigma_{V|U_{pj}}}\right)\right) \prod_{p=n_i+1}^{N_i} \Phi\left(\frac{q - \omega_j X_j'\beta - (1-\omega_j)\alpha_j^*}{\sigma_V}\right),$$

where $f(U, V)$ is the joint density of U_{pj} and V, V_{pj} is a function of the utility that *ex ante* peer p is expected to gain: $V_{pj} = \omega_j X_j'\beta + (1-\omega_j)s_{pj} - u_{p2}$, and ϕ is the standard normal density.[38] The maximum likelihood estimator in Week 2 is unbiased and approximately normal, $S_{ij2} \sim N\left(\alpha_j^*, \frac{1}{b_{i2}}\right)$.[39] Its precision is the Fisher information:

$$b_{i2} \equiv -E\left[\frac{\partial^2 \ln L_{ij2}}{\partial \alpha_j^{*2}}\right] = dn_i + (N_i - n_i)\frac{\phi(c)}{\Phi(c)}\left(c + \frac{\phi(c)}{\Phi(c)}\right)\left(\frac{1-\omega_j}{\sigma_V}\right)^2. \tag{A.2}$$

The precision of the maximum likelihood estimator varies across individuals because different individuals have different numbers of peers, N_i, and receive different numbers of feedbacks, n_i.[40]

APPENDIX 2

(1) No repeated purchases. Here, I consider a case in which most consumers do not go to the same movie twice so that ticket sales for a given movie to decline over time. While the probability that the representative consumer sees a movie in Week 1 is the same P_1 defined in equation (7), the probability for subsequent weeks changes. Consider first the case where there is no social learning. The probability that the representative consumer sees the movie in Week 2 is now the joint probability that her expected utility at $t = 2$ is higher than her cost and her expected utility at $t = 1$ is lower than her cost: $P_2 = \text{Prob}(E_1[U_{ij} \mid X_j'\beta, s_{ij}] < q_{i1}$ and $E_2[U_{ij} \mid X_j'\beta, s_{ij}] > q_{i2})$. It is clear that $P_2 < P_1$. Intuitively, many of the consumers who expect to like a given movie watch it during the opening weekend. Those left are less likely to expect to like the movie, so that attendance in the second week is weaker than in the first.

37. Note that in this set-up, there is information not only in feedback from peers who have seen the movie but also in the fact that some peers have decided not to see the movie. Since every individual receives an independent signal on movie quality, the fact that some of her peers have decided not to see the movie provides valuable additional information to consumer i.

38. The term σ_V is equal to $\sqrt{(1-\omega_j)^2\left(\frac{1}{d} + \frac{1}{k_j}\right) + \frac{1}{r}}$ and $\sigma_{V|U_{pj}} = \sqrt{(1-\omega_j)^2\left(\frac{1}{k_j}\right) + \frac{1}{r}}$.

39. The maximum likelihood estimate is the value of α_j^* that solves $\alpha_j^* = \frac{1}{n_i}\sum_{p=1}^{n_i} U_{pj} - \frac{N_i - n_i}{n_i}\frac{(1-\omega_j)}{d\sigma_V}\frac{\phi\left(\frac{q-\omega_j X_j'\beta-(1-\omega_j)\alpha_j^*}{\sigma_V}\right)}{\Phi\left(\frac{q-\omega_j X_j'\beta-(1-\omega_j)\alpha_j^*}{\sigma_V}\right)}$. Although this expression cannot be solved analytically, it is clear that the maximum likelihood estimate is less than the simple average of the utilities U_{pj} reported by peers who saw the movie. It is dampened by a "selection-correcting" term that increases with the fraction of peers who did not see the movie.

40. The term c is equal to $(q - \omega_j X_j'\beta - (1-\omega_j)\alpha_j^*)/\sigma_V$. Since $E[x \mid x < c] = \frac{\phi(c)}{\Phi(c)}$ for a standard normal variable x, it is clear that $c > \frac{\phi(c)}{\Phi(c)}$, b_{i2} is always positive, and the likelihood function is globally concave.

The key point here is that, while all movies exhibit a decline in sales over time, this decline is more pronounced for movies that experience strong sales in the first weekend. In particular, under these assumptions, it is possible to show that $\frac{\partial^2 P_t}{\partial t \partial \alpha_j^*} < 0$. A strong performance in Week 1 reduces the base of potential customers in Week 2. The effect of this intertemporal substitution is that the decline in sales over time is accelerated compared to the case of a movie that has an average performance in Week 1, although the total number remains higher because the intercept is higher. The opposite is true if a movie is worse than expected. See the bottom left panel of Figure 1.

With social learning, sales dynamics depend on the strength of the social learning effect:

$$\begin{aligned} P_2 &= \text{Prob}(E_2[U_{ij} \mid X'_j\beta, s_{ij}, S_{ij2}] > q_{i2} \text{ and } E_1[U_{ij} \mid X'_j\beta, s_{ij}] < q_{i1}) \\ &= \text{Prob}((\omega_{j22}(\nu_{ij} + \varepsilon_{ij}) + \omega_{j32}(S_{ij2} - \alpha_j^*) - u_{i2}) > (q - (1 - \omega_{j12})\alpha_j^* - \omega_{j12}X'_j\beta) \\ &\quad \text{and } (\omega_{j21}(\nu_{ij} + \varepsilon_{ij}) + \omega_{j31}(S_{ij1} - \alpha_j^*) - u_{i1}) < (q - (1 - \omega_{j11})\alpha_j^* - \omega_{j11}X'_j\beta)). \end{aligned}$$

If social learning is weak, the dynamics of sales will look qualitatively similar to the ones in the bottom left panel of Figure 1, although the slope of the movie characterized by a positive (negative) surprise is less (more) negative. But if social learning is strong enough, the dynamics of sales will look like the ones in the bottom right panel of Figure 1, where the slope of the movie characterized by a positive (negative) surprise is less (more) negative than the slope of the average movie.

(2) Option value. In my setting, I follow Bikhchandani, Hirshleifer and Ivo (1992) and model the timing of purchase as exogenous. This assumption rules out the possibility that consumers might want to wait for uncertainty to be resolved before making a decision. In the case of the latter possibility, consumers would have an expected value of waiting to decide, as in the Dixit and Pindyck (1994) model of waiting to invest. This would give rise to an option value associated with waiting. Like in the myopic case described above, a consumer in this setting decides to see the movie in Week 1 only if her private signal on quality is high enough relative to the opportunity cost of time. However, the signal that triggers consumption in the option value case is higher than its equivalent in the myopic case because waiting generates information and therefore has value. This implies a lower probability of going to see the movie in Week 1.

If ε_{ij} and q_{it} remain independent of all individual and movie characteristics and individuals take their peers' timing as given, the model generates the same set of implications. While decisions are more prudent in the strategic case than in the myopic case, the timing of purchase remains determined by the realization of the signal and of q_{it} and thus remains unsystematic. Therefore, information diffusion follows similar dynamics to those described above.[41]

Acknowledgment. I thank the editor, three anonymous referees, Jerome Adda, David Ahn, Raj Chetty, Roger Gordon, Bryan Graham, David Levine, Robert McMillan, Stefano Della Vigna, and seminar participants at Bank of Italy, Berkeley, Cornell, Chicago GSB, Nuremberg, LSE, Crest-Paris, Regensburg, San Diego, UCL, and Zurich for helpful comments. I am grateful to Stefano Della Vigna and Phillip Leslie for sharing their box-office data. Gregorio Caetano, Mariana Carrera, Ashley Langer, Zach Liscow, David Klein, and Erin Metcalf provided excellent research assistance.

REFERENCES

BANERJEE, A. (1992), "A Simple Model of Herd Behavior", *The Quarterly Journal of Economics*, **107**, 797–817.

BASS, F. (1969), "A New Product Growth for Model Consumer Durables", *Management Science*, **15**, 215–227.

BECKER, G. S. (1991), "A Note on Restaurant Pricing and Other Examples of Social Influences on Price", *Journal of Political Economy*, **99**, 1109–1116.

BIKHCHANDANI, S., HIRSHLEIFER, D. and WELCH, I. (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", *Journal of Political Economy*, **100**, 992–1026.

BIKHCHANDANI, S., HIRSHLEIFER, D. and WELCH, I. (1998), "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades", *Journal of Economic Perspectives*, **12**, 151–170.

CAI, H., CHEN, Y. and FANG, H. (2007), *Observational Learning: Evidence from a Randomized Natural Field Experiment* (Mimeo, Duke University).

CHEVALIER, J. and MAYZLIN, D. "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research*, forthcoming.

DAHL, G. and DELLAVIGNA, S. (2009), "Does Movie Violence Increase Violent Crime?", *Quarterly Journal of Economics*, **124**, 677–734.

DIXIT, A. K. and PINDYCK, R. S. (1994), *Investment under Uncertainty*, (Princeton, Princeton University Press).

EINAV, L. (2007), "Seasonality in the U.S. Motion Picture Industry", *Rand Journal of Economics*, **38**, 100–120.

41. A more complicated scenario arises if timing of purchase strategically depends on peers' timing. This could happen, *e.g.* if some individuals wait for their friends to go see the movie in order to have a more precise estimate of their signal, and their peers wait for the same reason. This scenario has different implications.

FRIEND, T. (2009), "The Cobra", The New Yorker, 19 January, 2009.

GLAESER, E. L., SACERDOTE, B. I., and SCHEINKMAN, J. A. (2003), "The Social Multiplier," *Journal of the European Economic Association*, **1**, 345–353.

GOETTLER, R. and LESLIE, P. (2005), "Cofinancing to Manage Risk in the Motion Picture Industry", *Journal of Economics and Management Strategy*, **14**, 231–261.

GRILICHES, Z. (1957), "Hybrid Corn: An Exploration in the Economics of Technical Change", *Econometrics*, **25**, 200–220.

KNIGHT, B. and SCHIFF, N. (2007), "Momentum and Social Learning in Presidential Primaries" (Working Paper No. 13637, NBER).

MORETTI. (2008), "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales" (Working Paper No. 13832, NBER).

REINSTEIN, D. A. and SNYDER, C. M. (2005), "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics -super-*", *Journal of Industrial Economics*, **53**, 27–51.

SALGANIK, M. J., DODDS, P. S. and WATTS, D. J. (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science*, **311**, 854–856.

SORENSEN, A. "Social Learning and Health Plan Choice", *RAND Journal of Economics*, forthcoming.

VISE, D. (2005), *The Google Story* (New York: Delacorte Press).

WELCH, I. (1992), "Sequential Sales, Learning, and Cascades", *Journal of Finance*, **47**, 695–732.