**Social Learning for Experiential Product Purchases: The Impact of Homophily and Balance**

Jae Young Lee and Raghuram Iyengar*

June 2013

* Jae Young Lee is a doctoral candidate (Email: jaely@wharton.upenn.edu) and Raghuram Iyengar is Associate Professor at the Wharton School, University of Pennsylvania (Email: riyengar@wharton.upenn.edu). The authors benefited from comments by David Bell, Eric Bradlow, Rom Schrift, Christophe Van den Bulte, and audience members at the University of Houston Doctoral Consortium and the Fall 2012 Wharton School Marketing Seminar series.

**Social Learning for Experiential Product Purchases: The Impact of Homophily and Balance**


## Abstract

Consumers often turn to social contacts prior to product purchases. We study how interpersonal relationships among consumers and their contacts impact information search, learning and purchase of experiential products. There are two important features of our framework. First, the learning model is purposive with active search for a planned purchase and flexibly accommodates aspects of interpersonal relationships among information recipients and sources. Second, consumers' search and purchase decisions are interrelated, but temporally separated, thus allowing us to assess the impact of social relationships on each decision. The model is estimated using incentive compatible stated choice experiments and there are two key insights. First, consumers prefer to gather information from similar others (i.e., homophily) due to the greater informational benefit from similar others rather than the convenience of collecting information. Second, triadic balance in a social system has a nuanced effect. While people understand that informational benefits are greater under an imbalanced relationship, the cost of information seeking is also higher. Consumers search less under an imbalanced social system as compared to when it is balanced and yet the lower amount of search still leads to greater purchase likelihood. Our findings have relevance for Internet retailers offering social evaluations for products.



KEYWORDS: Social Learning; Consumer Search; Homophily; Triadic Balance; Multivariate Bayesian Learning.

## 1. Introduction

Consumers often turn to friends or other social contacts before making purchases as mundane as the movie to watch on a Friday night or decisions as critical as where to have a wedding. Social sources are especially useful for gathering information about experiential product attributes, which cannot be fully evaluated prior to direct experience (e.g., how much I will like a restaurant), rather than search attributes, which can be verified pre-purchase (e.g., price of an entree in a restaurant). As an example, the Internet portal Yelp.com provides easy access to information about location, price, and menu of restaurants (search attributes). It also offers aggregate popularity for restaurants but the evaluations are typically from people who consumers do not know and hence predicting their own liking for a restaurant is difficult. Therefore, consumers frequently gather feedback from others whom they know and have shared common experiences with. The ubiquity of Internet-based communication channels has increased the viability of such social search (Godes et al. 2005): Several Internet retailers such as goodreads.com (books), opentable.com (restaurants), netflix.com (movie / DVD) and snoox.com (entertainment) have begun providing friends' reviews to their customers.

Social information is distinct from the information via other sources (e.g., targeted marketing effort) for at least three reasons. First, sources of social information (i.e., network contacts) may have systematically different product tastes as compared to a focal consumer. Second, social sources themselves may have similar or dissimilar tastes with each other. Finally, it is likely that the features of interpersonal relationships among information recipients and sources such as *homophily* (the tendency of individuals to associate with similar others) and *balance* (the overall consistency of preferences in a social system) influence learning.

We study whether and how the similarity of preference in a social system influences the way that consumers actively gather information, learn from it and make purchase decisions. There are two important aspects of our model. First, the model extends standard multivariate Bayesian learning models (e.g. Erdem 1998; Winkler 1981) by accommodating features of interpersonal relationships to better describe social learning. Second, consumers' search and purchase decisions are modeled as interrelated but temporally separated (e.g., Erdem et al. 2005) thus allowing us to assess the impact of the similarity of preferences with social sources on each decision.

Our data for the empirical application are from an incentive compatible stated choice experiment where consumers make purchase decisions for individual music tracks while having access to others' evaluations. For addressing research questions regarding the underlying mechanisms for social learning, such data have several advantages over observational data (e.g., Hartmann et al. 2008; Narayan et al. 2011). First, we manipulate the similarity of preference, which is usually confounded with interpersonal affect or frequency of interaction in observational data. Second, the experimental data control for unobserved confounds for identifying social influence such as endogenous group formation. Third, there was no possibility of passive social learning (from exogenous social information), awareness diffusion, or normative pressure in our study. Finally, we also manipulated the content and availability of social information, which is difficult to observe in secondary data. Thus, we can isolate the effect of the similarity structure on consumers' decisions while ensuring external validity with an incentive-aligned experiment.

Our results provide two key insights regarding social learning. First, there is evidence of homophily as consumers prefer to gather information from similar others. Furthermore, we pinpoint the mechanism behind this phenomenon. Our results suggest that homophily is driven by greater informational benefit from similar others (Suls et al 2000) rather than the comfort of collecting such information (Carley 1991). Second, balance of a social system has a more nuanced effect on social learning. On the one hand, people appear to understand that informational benefit is greater under an imbalanced social system (Goethels and Nelson 1973; Levine et al. 1993) than a balanced one. On the other hand, people tend to have greater discomfort under an imbalanced system (Heider 1946; Meyers-Levy and Tybout 1989), so the cost of information-seeking is higher. Our results suggest that people tend to search less under an imbalanced system, as compared to a balanced one, due to search costs. However, the lower amount of search under imbalance still leads to greater informational benefit as compared to a balanced condition.

Our findings are relevant for many Internet retailers who offer social reviews to increase product sales (e.g., Flixster, Opentable, Goodreads, etc). Our policy experiments show that firms can increase their expected profit by providing customers with product reviews from others with whom they have an imbalanced relationship. The benefit of imbalance on expected profit is enhanced when firms diversify the share of offered reviews across different social groups that their customers may have. In addition, although consumers may be willing to access more reviews under a balanced social system due to lower cost, firms should still provide fewer reviews. This is because while an additional review under a balanced system may be easier to read for consumers, it does not sufficiently increase the likelihood of purchase.

The remainder of the paper is organized as follows. Section 2 describes relevant prior research. Section 3 describes our research setting, and we develop a model for social learning and the two inter-related decisions (search and purchase) in Section 4. Section 5 outlines the empirical application of our model to the data collected from an incentive compatible stated choice experiment. Section 6 reports the findings, and Section 7 discusses managerial implications. Section 8 concludes with contributions and key results.

## 2. Literature Review

Our study is at the intersection of two streams of research, consumer learning (i.e., information partly resolves the uncertainty about experiential attributes; see Ching et al. 2012 for an overview) and search (i.e., consumers actively seek information; see Ratchford 2010 for an overview). In addition, we focus on social contacts as sources, and investigate the role of social networks in search and learning.

Social learning has at least three features that are distinct from other types of consumer learning. First, a consumer could gather information from others who may have different tastes compared to his own. This is unlike learning from own experience (e.g., Erdem and Keane 1996) or targeted detailing (e.g., Narayanan and Manchanda 2009). Past studies on social learning (Roberts and Urban 1988; Erdem et al. 2005; Zhao et al. 2012) have not addressed this issue adequately and assumed that social reviews provide unbiased signals for consumers' evaluation.

Second, social contacts may have similar (or dissimilar) tastes, which creates interdependency among sources. There are several studies which have proposed a standard multivariate Bayesian learning model (SMBL) to incorporate the interdependency of information sources (Erdem 1998; Winkler 1981). However, SMBL model has two major restrictions that limit its applicability in a social context (described briefly here but see Section 5.2 and Appendix 2) – (1) informational benefit from others with similar and dissimilar preferences is equivalent, and (2) there is always greater informational benefit when a social network has imbalanced preferences as compared to balanced ones.

Third, learning is likely to be driven by characteristics of social relationships among information receivers and providers. There are studies that have tried to accommodate behavioral aspects of learning such as forgetting (Mehta et al. 2004), the salience of recent signals (Camacho et al. 2011), and the valence of signals (Zhao et al. 2011). However, a learning model that accommodates the interpersonal aspects of relationships, which is crucial for social learning, has not been well developed.

We propose an extended multivariate Bayesian learning (EMBL) model which incorporates all three features of social learning. In our model, consumers gather information from social contacts whose preferences are allowed to be systematically different from their own evaluation; consumers can still infer their evaluation given the similarity of preference with their social contacts (Goethals and Darley 1977; Yaniv et al. 2011). The model also allows for the interdependency among social information sources. Most importantly, EMBL model flexibly accommodates the interpersonal aspects of social learning. We focus on two features of social relationships – homophily and balance, which are described next.

Homophily is the tendency of individuals to associate with similar others (see McPherson et al. 2001 for a review). Homophily should matter for social learning for two reasons. First, people may believe that the personal experience of similar others as compared to that of dissimilar others provides a better forecast (i.e., informational benefit) of one's own future experience (Suls et al. 2000). Note that homophily driven by informational benefit cannot be accommodated in an SMBL model as positive feedback from similar others and negative feedback from dissimilar others are equivalent (see Section 5.2 and Appendix 2). Second, people may have greater discomfort (mental cost) interacting with dissimilar others than similar others (Carley 1991). We test whether homophily impacts consumers' decisions, and further identify which of the two potential drivers (or both) is responsible for any observed effect.

A social system is balanced if the valence of preference similarity multiplies out to be positive (Cartwright and Harary 1956; Heider 1946). There are at least two reasons why balance in preferences within a social system should matter for social learning. First, balance in a social relationship may affect the informational benefit that consumers can obtain from social contacts. On the one hand, exposure to different vantage points under an imbalanced relationship may increase information diagnosticity (Goethals and Nelson 1973; Levine et al. 1993). On the other hand, people may be more confident with their decisions under a balanced relationship as they are more likely to see congruent information (Tversky and Kahneman 1974). Note that SMBL does include a restrictive link between balance in social relationships and consumer learning: the information gathered from an imbalanced relationship is always more informative than a balanced relationship (see Section 5.2 and Appendix 2). Second, under imbalance, people may experience greater discomfort (Heider 1946) or need to put greater mental activity to comprehend inconsistent information (Meyers-Levy and Tybout 1989), and these may lead to greater search cost. We test whether and how balance within a social system impacts consumer learning and their decisions.

We focus on search for experiential attributes where (a) information sources may have product evaluations that are systematically different from those of a focal consumer and (b) feedback does not completely resolve uncertainty. The scenario is unlike that in the classic literature on consumer search (e.g., Stigler 1961; Weitzman 1979) in which the content of information hardly varies with information sources and all uncertainty is resolved after search. Therefore, we examine the choice of social information sources and the search amount from different sources as key aspects of search decision (Urban et al. 1993; Ratchford et al. 2003). Moreover, we *endogenize* consumer search on experiential attributes within a purchase model such that consumers search to make a more informed purchase decision (Erdem et al. 2005). The novelty of our approach is that we explicitly account for interpersonal features of social information sources, and assess its impact on "what source to search" and "how much to search". Next, the research setting provides details on the decisions that consumers make.

## 3. Problem Description

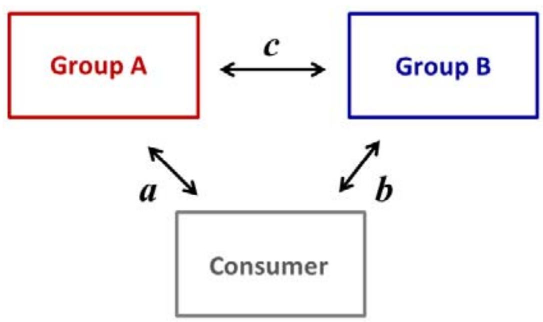
Prior to purchasing experiential products, consumers often turn to others who have experienced the product as they can provide evaluations. Such evaluations could be from (1) the general public (e.g., aggregate iTunes ratings for songs from typically anonymous reviewers) and from (2) social information sources (i.e., others who one knows). As aggregate information is easily accessible, it may be a part of prior knowledge for consumers before they decide to search for the latter. In this study, therefore, a search decision refers to how consumers collect product evaluations from social information sources.

Our interest is how the similarity of preference of a focal consumer with each social source, and between each pair of sources will affect his decision to search for product evaluations. This general problem is challenging and even intractable; if a consumer knows $n$ other people who have tried the product then, the similarity of preference with $n$ others (the direct connections) and the similarity of preference between $n(n-1)/2$ pairs will impact his search decision. Even with $n$=4 (a relatively small number) there are 10 different similarity parameters. Thus, to maintain the essence of the problem and make it tractable, we make the following assumption: a consumer's social sources can be categorized into two exogenous groups (groups A and B).

Given the context of two social groups, we define the search decision as how many people from each group does a consumer ask for product evaluation. Each individual evaluation is a "signal."

Figure 1 shows an example of a triad where the focal consumer and two social groups form the nodes and the link among any two nodes denotes the similarity of preference among them. In the figure, *a* (*b*) denotes the similarity of preference between a consumer and group A (group B), and *c* denotes the similarity of preference between the two social groups. These similarity measures are operationalized as standard correlation measure: the value gets closer to 1 (-1) as the positive (negative) association of preference between two nodes gets stronger. There is no association of preference between two nodes when the similarity measure between them is 0.

Figure 1. The Similarity of Preference in a Social Triad




We operationalize three measures summarizing the similarity of preference – relevance, redundancy, and balance – as follows. How certain the focal consumer will be about his own evaluation will depend on the strength of association between his preference and that of the social source - if a group of people, who have tastes that are very similar (dissimilar) to a focal consumer, give positive feedback about a product, he can infer that he may like (dislike) it (Goethals and Darley 1977). The strength of association is defined as *relevance*, and operationalized as the absolute value of similarity measure: the relevance of group A (B) is $|a|$

(|*b*|). As the reviews from one source are reasonable predictors for the reviews from the other source, the two social groups are redundant information sources: When the two social groups have very similar (dissimilar) tastes, one can expect that the signals from the two groups will be also very similar (dissimilar). Therefore, *redundancy* is defined as the strength of association in preference between the two social groups, and operationalized as $|c|$. Finally, the balance status is defined based on the sign of the product of preference similarities, *abc*: the triad is balanced (imbalanced) when $abc > 0$ ($abc < 0$).

We focus on the case where a consumer has a mature relationship with both social groups in a category and so knows the similarity of preference in the social triad (which consists of a consumer and the two social groups).[1] Also, a consumer makes a search decision before he observes any signal (but is knowledgeable regarding the aggregate product evaluation in the population), so a search decision in this study denotes simultaneous search.[2] Several real world contexts fit the research setting. For instance, suppose Ryan is deciding whether or not to dine in a particular restaurant on a Friday evening. As he shares a lot of common dining experiences with his friends, Ryan knows how similar his taste for restaurants is to his friends as well as how similar his friends' tastes are to each other. Ryan sends out multiple messages (e.g., group SMS messages) to his friends at the same time, and waits for their evaluations. The above scenario characterizes a situation where a consumer is time constrained and cannot sequentially decide to collect information from his friends.

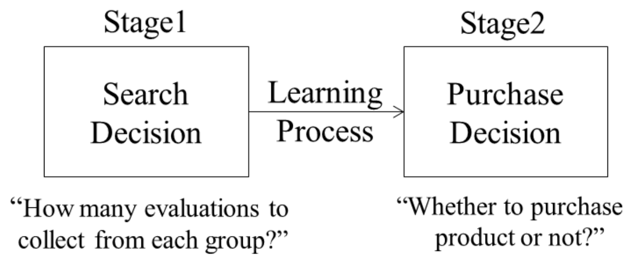
---

[1] There are other contexts where consumers are uncertain about their similarity of preference with others and learn about them over time. In addition, consumers' social contacts may be categorized into multiple (more than two) groups and these may not be exogenous. We discuss related extensions in future research.

[2] Morgan and Manning (1985) have found that either sequential or simultaneous search (or a combination of both) can be optimal for a consumer. More recently, De los Santos et al. (2012) estimated both simultaneous and sequential search models in the context of online search for experiential products, which is also the setting in our study, and found that a simultaneous search model fit the data better. Obviously, there are contexts in which consumers perform sequential search (Weitzman 1979). We discuss related extensions in future research.

Figure 2 summarizes the decision framework for a consumer. For each product, in Stage 1, a consumer decides how many evaluations to collect from each of the two social groups. After collecting signals from each group, a consumer updates his belief about how much he will like the product (i.e., his own product evaluation). Based on these updated beliefs, in Stage 2, a consumer decides whether or not to purchase the product.

Figure 2. Decision Framework for a Product




**4. Model**

For each product, a consumer makes two inter-connected, but temporally separated decisions: In stage 1 (at time $t_1$), a consumer decides how much information to acquire about product evaluation from other consumers (“search decision”) and in stage 2 (at time $t_2$), makes a binary decision of whether or not to purchase the product (“purchase decision”). Between the two stages, a consumer processes the information from search. In this section, we build a model for the entire process. Note that the framework proposed in this section is common for SMBL and EMBL models. We make a distinction between the two types of learning models in Section 5.

We assume that consumers are utility maximizers and both decisions are driven by the same utility function. Section 4.1 describes the utility function. In Section 4.2, we build a model for the search decision. We show how the belief distributions evolve after the search decision in Section 4.3, and finally model the purchase decision in Section 4.4.

### 4.1. Utility Specification

Let $x_j = (x_{j1}, x_{j2}, \ldots, x_{jm})$ denote a vector of $m$ attributes (including price) associated with product $j$. Let $\alpha_{ij}$ denote the attribute-based utility that consumer $i$ associates with product $j$. We assume $\alpha_{ij}$ has the following functional form:

$$\alpha_{ij} = \gamma_{i0} + \sum_{k=1}^{m} \gamma_{ik} x_{jk} \ . \tag{1}$$

Let $I_i(t)$ denote the information set of consumer $i$ at time $t$. For notation simplicity, we omit the product subscript $j$ in the information set. The information set at a particular time characterizes the state of the consumer and consists of all factors that affect current utility at time $t$ and any future utilities. In our setting, there are two time points, $t_1$ and $t_2$, and a consumer has a different information set at these two time points. Throughout this section, we will illustrate what the consumer knows at the two different time points.

We define a consumer $i$'s indirect utility from purchasing product $j$ at time $t$ using a constant absolute risk aversion (CARA) specification (Narayanan and Manchanda 2009; Zhao et al. 2012):

$$U_{ij}\left(I_i(t)\right) = \alpha_{ij} - \exp\left(-\beta_i R_{ij}^E\left(I_i(t)\right)\right) + \varepsilon_{ij}\left(I_i(t)\right), \qquad t = t_1 \text{ or } t_2. \tag{2}$$

The term $R_{ij}^E$ refers to consumer $i$'s rating (or evaluation) of product $j$ and is realized only *after* product experience. We use the term $R_{ij}^E\left(I_i(t)\right)$ to denote explicitly that consumer $i$'s knowledge about $R_{ij}^E$ at time $t$ depends on his information set $I_i(t)$. The parameter $\beta_i$ captures the effect of product evaluation on purchase utility. The error term $\varepsilon_{ij}\left(I_i(t)\right)$ is also dependent on the information set. We assume that at the time of search (Stage 1), the error term is stochastic to consumers while at the time of purchase (Stage 2), it is observable. The assumption implies that there is a temporal separation between the search and purchase decisions. Finally, the utility from not purchasing the product is set to 0.

### 4.2. Search Decision – Stage 1

To ease the explication of our search model, we first broadly explain how a consumer makes the purchase decision after conducting a specific amount of search. This discussion will illustrate the link between search and its impact on the purchase decision. Next, we specify the uncertain beliefs that consumers hold for stochastic variables at the time of search. Finally, we describe how consumers determine their optimal level of search given these prior beliefs.

**4.2.1 Link between Search and Purchase.** Suppose consumer *i* collects $n_{ij} = \left(n_{ij}^A, n_{ij}^B\right)$ individual evaluations from the two groups and the average of these collected evaluations is $\bar{s}_{ij} = \left(\bar{s}_{ij}^A, \bar{s}_{ij}^B\right)$. After conducting this search, consumer *i*'s information set at time $t_2, I_i(t_2)$, includes $n_{ij}$ and $\bar{s}_{ij}$. Let $f_i^{R^E}(R_{ij}^E \mid I_i(t_2))$ denote consumer *i*'s belief about his own evaluation given this information set.

As consumer *i* is uncertain about his own product evaluation, he will determine the expected utility of purchasing product *j* with respect to his beliefs about product evaluation. Consumer *i* will purchase product *j* if and only if it provides higher expected utility than not purchasing the product. Equivalently, consumer *i* will purchase product *j* if

$$E_{R_{ij}^E(I(t_2))}\left[U_{ij}\left(I_i(t_2)\right)\right] > 0, \tag{3}$$

where E[.] is the expectation operator. The consumer will not purchase a product *j* otherwise. Using the expression in Equation (2), the term $E_{R_{ij}^E(I(t_2))}\left[U_{ij}\left(I_i(t_2)\right)\right]$ can be expressed as:

$$E_{R_{ij}^E(I(t_2))}\left[U_{ij}\left(I_i(t_2)\right)\right] = E_{R_{ij}^E(I(t_2))}\left[\alpha_{ij} - \exp\left(-\beta_i R_{ij}^E(I(t_2))\right)\right] + \varepsilon_{ij}\left(I(t_2)\right), \tag{4}$$

$$= u_{ij}\left(I_i(t_2)\right) + \varepsilon_{ij}\left(I(t_2)\right),$$

where $u_{ij}\left(I_i(t_2)\right)$ denotes the systematic component of the expected utility of purchase. Note that the random component of utility, $\varepsilon_{ij}$ , is observable to consumers at Stage 2.

The above description emphasizes that a consumer's decision of whether or not to purchase a product depends on his earlier search decision as the number and the type of signals ($n_{ij}$ and $\overline{s}_{ij}$ ) alter his information set at the time of purchase, $I_i(t_2)$ .

**4.2.2. Consumer Beliefs.** In this section, we elaborate on consumers' beliefs about the relevant stochastic variables given the search decision, $n_{ij}$ , and the information set at the search stage, $I(t_1)$ . We specify beliefs they hold about (1) own evaluation prior to search, $R_{ij}^E\left(I(t_1)\right)$, (2) signals to be observed, $\overline{s}_{ij}\left(I(t_1), n_{ij}\right)$, and (3) random component of utility, $\varepsilon_{ij}\left(I(t_1)\right)$.

*Prior Belief about Evaluations***.** We make the following assumptions to specify consumers' belief about their own product evaluation prior to search. As emphasized, consumer *i* is uncertain about his own evaluation $\left(R_{ij}^E\right)$ for a product *j*. Likewise, we assume that he is also uncertain about the average evaluations in the two social groups $\left(R_{ij}^A, R_{ij}^B\right)$. Therefore, a consumer has uncertain beliefs about vector $R_{ij} = \left(R_{ij}^E, R_{ij}^A, R_{ij}^B\right)$, which is represented by the distribution, $f_i^R\left(R_{ij} \mid I_i(t_1)\right)$.

As explained earlier, a consumer knows the aggregate distribution of product evaluation when he makes a search decision ($t_1$). We assume that the aggregate distribution is from a sufficiently large number of people such that it well represents the population distribution. At the search stage, therefore, a consumer believes that his own rating $\left(R_{ij}^E\right)$ and the average ratings in the two social groups $\left(R_{ij}^{A}, R_{ij}^{B}\right)$ are drawn from this population distribution. We assume that the population distribution is normally distributed with mean $R_{ij}^{0}$ and variance $\tau_{ij}^{2}$.

A consumer also knows that how similar (or dissimilar) his preferences are to those of each social group $\left(a_{ij}, b_{ij}\right)$, and how similar (or dissimilar) the preferences of the two social groups are $\left(c_{ij}\right)$: We assume that similarity of preference in the social triad, $\rho_{ij} = \left(a_{ij}, b_{ij}, c_{ij}\right)$, is known to a consumer from his past experience in a category *ex-ante*. Therefore, a consumer believes that the vector $R_{ij}$ is a multivariate sample from the population distribution. Thus, we express $f_i^R\left(R_{ij} \mid I_i(t_1)\right)$ as:

$$f_i^R\left(R_{ij} \mid I_i(t_1)\right) = N\left(\begin{bmatrix} R_{ij}^0 \\ R_{ij}^0 \\ R_{ij}^0 \end{bmatrix}, \tau_{ij}^2 \times \begin{bmatrix} 1 & a_{ij} & b_{ij} \\ a_{ij} & 1 & c_{ij} \\ b_{ij} & c_{ij} & 1 \end{bmatrix}\right). \tag{5}$$

Note that as a covariance matrix should be positive definite, there are restrictions on the values that similarity of preference $\left(\rho_{ij}\right)$ can take (see Appendix 1). Given Equation 5, a consumer *i*'s belief about his own rating, $f_i^{R^E}\left(R_{ij}^E \mid I_i(t_1)\right)$, is obtained from the marginal distribution:

$$f_i^{R^E}\left(R_{ij}^E \mid I_i(t_1)\right) = N\left(R_{ij}^0, \tau_{ij}^2\right). \tag{6}$$

*Signal Distribution***.** To specify the belief about evaluation signals that a consumer may receive by searching, we make the following assumptions. First, a consumer understands that there is variation of individual evaluations (signals) within each group and any signal is distributed around the average evaluation within a group. Second, a consumer believes that a signal from each group follows *i.i.d.* normal distribution with an unknown average evaluation within a group ( $R_{ij}^{A}\left(I_i(t_1)\right)$ or $R_{ij}^{B}\left(I_i(t_1)\right)$) and standard deviation ( $\sigma_{ij}^{A}$ or $\sigma_{ij}^{B}$ ).

Given these assumptions, the belief about sample average of $n_{ij} = \left(n_{ij}^{A}, n_{ij}^{B}\right)$ signals, $f_i^{\bar{s}}\left(\bar{s}_{ij} \mid I_i(t_1), n_{ij}\right)$, can be expressed as:

$$f_i^{\bar{s}}\left(\bar{s}_{ij} \mid I_i(t_1), n_{ij}\right) = N\left(\begin{bmatrix} R_{ij}^{0} \\ R_{ij}^{0} \end{bmatrix}, \begin{bmatrix} \tau_{ij}^{2} + \left(\sigma_{ij}^{A}\right)^{2} / n_{ij}^{A} & c_{ij}\tau_{ij}^{2} \\ c_{ij}\tau_{ij}^{2} & \tau_{ij}^{2} + \left(\sigma_{ij}^{B}\right)^{2} / n_{ij}^{B} \end{bmatrix}\right), \tag{7}$$

where the expression is obtained by combining the uncertainty about the average among collected signals from each group and the uncertainty about average evaluation in each group. We assume that each social group consists of sufficiently large number of people and a consumer can collect any number of signals.

*Utility Error***.** The utility error is stochastic from consumers' perspective in the search stage. A consumer believes that utility error is normally distributed with mean 0 and unit variance, i.e., $f_i^{\varepsilon}(\varepsilon_{ij} \mid I_i(t_1)) = N(0,1)$. Next, we show how a consumer, who is fully aware of the link between the two decisions and the uncertainty about $I_i(t_2)$, will optimally conduct the search decision.

**4.2.3 Optimal level of Search.** The search for product evaluations from others is useful as it helps consumers to make a more informed purchase decision. For consumer *i*, let $k_{ij}^{A}$ and $k_{ij}^{B}$ denote the search cost for obtaining an evaluation from the social groups A and B, respectively. Such cost could be due to the hassle of collecting information or cost of processing information. The cost can differ by group, and the consumer knows this cost.

Given the information set of consumer *i* at $t_1$, $I_i(t_1)$, the utility from search $U_{ij}^{S}(.)$ for a specific amount of search $n_{ij}$ is as follows.

$$U_{ij}^{S}\left(I_i(t_1),\ n_{ij}\right)=\begin{cases} U_{ij}\left(I_i(t_2)\mid I_i(t_1),n_{ij}\right)-k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}+\xi_{n_{ij}}, & \text{when product } j \text{ is purchased,} \\ -k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}+\xi_{n_{ij}}, & \text{when product } j \text{ is not purchased.} \end{cases} \tag{8}$$

Here the term $\xi_{n_{ij}}$ is known to the consumer. The term can be interpreted as a fixed cost of gathering $n_{ij}$ evaluations (De los Santos, Hortaçsu and Wildenbeest 2013). We use the term $U_{ij}\left(I_i(t_2)\mid I_i(t_1),n_{ij}\right)$ to denote explicitly that the consumer *i*'s utility is based on how his information set at $t_2$ will change due to his information at $t_1$ and the amount of search he decides to engage in.

It is worthwhile to discuss the term $U_{ij}\left(I_i(t_2)\mid I_i(t_1),n_{ij}\right)$ in more detail. It contains two key components that are uncertain to consumers at time $t_1$. First, the utility error is stochastic, and consumer believes that it follows $f_i^{\varepsilon}(\varepsilon_{ij}\mid I_i(t_1))$, as shown in Section 4.2.2. Second, the consumer is yet to observe any product evaluations, and his uncertain belief follows $f_i^{\bar{s}}(\bar{s}_{ij}\mid I_i(t_1),n_{ij})$, as also shown in Section 4.2.2. The latter is important as it indicates that a consumer does not know what beliefs he will hold about his own product evaluation, $f_i^{R^E}(R_{ij}^{E}\mid I(t_2))$, at the time of purchase. Thus, a consumer *i*'s expected utility from search for product *j*'s evaluations is:

$$E\left[U_{ij}^{S}\left(I_i(t_1),\ n_{ij}\right)\right]=E_{I_i(t_2)\mid I_i(t_1)}\left[E_{R_{ij}^{E}(I(t_2))}\left[U_{ij}\left(I_i(t_2)\mid I_i(t_1),n_{ij}\right)\right]\right]-k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}+\xi_{n_{ij}}\ . \tag{9}$$

The above equation implies that expected search utility for a consumer equals the expected purchase utility (with respect to uncertain belief about own evaluation after search) after integrating over all possible $I(t_2)$ he may have given $I(t_1)$. Given Equations 3 and 4, we can rewrite Equation 9 as:

$$E\left[U_{ij}^{S}\left(I_i(t_1),\ n_{ij}\right)\right]=E_{I_i(t_2)|I_i(t_1)}\left[1_{I_i(t_2)}\times\left(u_{ij}\left(I_i(t_2)\right)+\varepsilon_{ij}\left(I(t_2)\right)\right)\right]-k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}+\xi_{n_{ij}}\ , \quad (10)$$

where $1_{I_i(t_2)}$ denotes an indicator which is 1 if a consumer *i* purchases a product *j*, and 0 otherwise. A purchase decision is made based on Equation 3, so it is a function of $u_{ij}\left(I_i(t_2)\right)$ and $\varepsilon_{ij}\left(I(t_2)\right)$.

The expectation operator $E_{I_i(t_2)|I_i(t_1)}[\cdot]$, can be decomposed into the expectation over signals that a consumer may receive, $E_{\bar{s}_{ij}(I_i(t_2)|I_i(t_1))}[\cdot]$, and the expectation over the utility errors, $E_{\varepsilon_{ij}(I_i(t_2)|I_i(t_1))}[\cdot]$. Equation 10 can be rewritten by using the properties of conditional expectation and normality of utility error as:

$$E\left[U_{ij}^{S}\left(I_i(t_1),\ n_{ij}\right)\right]=E_{\bar{s}_{ij}(I_i(t_2)|I_i(t_1))}\left[E_{\varepsilon_{ij}(I_i(t_2)|I_i(t_1))}\left[1_{I_i(t_2)}\times\left(u_{ij}\left(I_i(t_2)\right)+\varepsilon_{ij}\left(I(t_2)\right)\right)\right]\right] \quad (11)$$
$$-k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}+\xi_{n_{ij}}\ ,$$

$$=E_{\bar{s}_{ij}(I_i(t_2)|I_i(t_1))}\left[\Pr\left[u_{ij}\left(I_i(t_2)\right)+\varepsilon_{ij}\left(I(t_1)\right)>0\right]\right.$$
$$\left.\times E\left[u_{ij}\left(I_i(t_2)\right)+\varepsilon_{ij}\left(I(t_1)\right)\mid u_{ij}\left(I_i(t_2)\right)+\varepsilon_{ij}\left(I(t_1)\right)>0\right]\right]-k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}+\xi_{n_{ij}}\ ,$$

$$=\underbrace{E_{\bar{s}_{ij}(I_i(t_2)|I_i(t_1))}\left[\Phi\left[u_{ij}\left(I_i(t_2)\right)\right]u_{ij}\left(I_i(t_2)\right)+\phi\left[u_{ij}\left(I_i(t_2)\right)\right]\right]}_{\text{Informational Benefit from Search}}\underbrace{-k_{ij}^{A}n_{ij}^{A}-k_{ij}^{B}n_{ij}^{B}}_{\text{Cost of Search}}+\xi_{n_{ij}}\ ,$$

$$=v^{b}\left(n_{ij}\mid I_i\left(t_1\right)\right)-v^{c}\left(n_{ij}\right)+\xi_{n_{ij}}\ ,$$

where $v^{b}\left(n_{ij}\mid I_i\left(t_1\right)\right)$ denotes the expected informational benefit from search, and $v^{c}\left(n_{ij}\right)$ denotes the cost of search.

We assume that the consumer evaluates the expected utility associated with each level of search. He then chooses the level ($n_{ij}^{*}$) that maximizes the expected utility from search. Thus,

$$n_{ij}^{*} = \arg\max_{n_{ij}} E\left[U_{ij}^{S}\left(I_i(t_1),\ n_{ij}\right)\right]. \tag{12}$$

**4.3. Learning Process.**

After collecting signals from each group, consumers update their belief about not only their own product evaluation $\left(R_{ij}^{E}\right)$ but also the average evaluations in the two social groups $\left(R_{ij}^{A}, R_{ij}^{B}\right)$. We assume that consumers update their beliefs about all evaluations according to Bayes rule. Given the prior and the signal distribution specified in Section 4.2.2, the learning process follows multivariate Bayesian learning. As a result, we obtain the posterior belief about all three evaluations, $f_i^{R}\left(R_{ij} \mid I_i(t_2)\right)$, which follows a multivariate normal distribution. Thus, we can obtain a consumer's posterior beliefs about their own product evaluation, $f_i^{R^E}\left(R_{ij}^{E} \mid I_i(t_2)\right)$, as a marginal distribution of $f_i^{R}\left(R_{ij} \mid I_i(t_2)\right)$:

$$f_i^{R^E}\left(R_{ij}^{E} \mid I_i(t_2)\right) = N\left(PM_{ij}\left(n_{ij}, \overline{s}_{ij}\right), PV_{ij}\left(n_{ij}\right)\right), \tag{13}$$

where $PM_{ij}\left(n_{ij}, \overline{s}_{ij}\right) = \omega_{ij}^{0} R_{ij}^{0} + \omega_{ij}^{A} \overline{s}_{ij}^{A} + \omega_{ij}^{B} \overline{s}_{ij}^{B}$,

$$\omega_{ij}^{A} = \frac{\tau_{ij}^{2} n_{ij}^{A}\left(a_{ij}\left(\left(\sigma_{ij}^{B}\right)^{2} + \tau_{ij}^{2} n_{ij}^{B}\right) - b_{ij} c_{ij} \tau_{ij}^{2} n_{ij}^{B}\right)}{\left(\left(\sigma_{ij}^{A}\right)^{2} + \tau_{ij}^{2} n_{ij}^{A}\right)\left(\left(\sigma_{ij}^{B}\right)^{2} + \tau_{ij}^{2} n_{ij}^{B}\right) - c_{ij}^{2} \tau_{ij}^{4} n_{ij}^{A} n_{ij}^{B}},$$

$$\omega_{ij}^{B} = \frac{\tau_{ij}^{2} n_{ij}^{B}\left(b_{ij}\left(\left(\sigma_{ij}^{A}\right)^{2} + \tau_{ij}^{2} n_{ij}^{A}\right) - a_{ij} c_{ij} \tau_{ij}^{2} n_{ij}^{A}\right)}{\left(\left(\sigma_{ij}^{A}\right)^{2} + \tau_{ij}^{2} n_{ij}^{A}\right)\left(\left(\sigma_{ij}^{B}\right)^{2} + \tau_{ij}^{2} n_{ij}^{B}\right) - c_{ij}^{2} \tau_{ij}^{4} n_{ij}^{A} n_{ij}^{B}},$$

$$\omega_{ij}^{0} = 1 - \omega_{ij}^{A} - \omega_{ij}^{B},$$

and $PV_{ij}\left(n_{ij}\right) = \tau_{ij}^{2} \times \dfrac{\left(\tau_{ij}^{2} n_{ij}^{A}\left(1 - a_{ij}^{2}\right) + \left(\sigma_{ij}^{A}\right)^{2}\right)\left(\tau_{ij}^{2} n_{ij}^{B}\left(1 - b_{ij}^{2}\right) + \left(\sigma_{ij}^{B}\right)^{2}\right) - \tau_{ij}^{4} n_{ij}^{A} n_{ij}^{B}\left(a_{ij} b_{ij} - c_{ij}\right)^{2}}{\left(\left(\sigma_{ij}^{A}\right)^{2} + \tau_{ij}^{2} n_{ij}^{A}\right)\left(\left(\sigma_{ij}^{B}\right)^{2} + \tau_{ij}^{2} n_{ij}^{B}\right) - c_{ij}^{2} \tau_{ij}^{4} n_{ij}^{A} n_{ij}^{B}}$.

Note that the Bayesian learning process is common to both SMBL and EMBL. The difference between the two models is whether update is based on objective attributes of the decision context (SMBL) or subjective attributes (EMBL). We discuss the difference in Section 5.2.

**4.4. Purchase Decision – Stage 2**

In the purchase stage, a consumer determines whether or not to purchase the product. Given the normality of consumer's beliefs about their own evaluation at the time of purchase, the expected utility from purchasing product $j$ is:

$$\begin{aligned} E_{R_{ij}^{E}(I(t_2))}\left[U_{ij}\left(I_i(t_2)\right)\right] &= \alpha_{ij} - \exp\left(-\beta_i PM_{ij}\left(n_{ij}, \overline{s}_{ij}\right) + \frac{\beta_i^2}{2} PV\left(n_{ij}\right)\right) + \varepsilon_{ij}\left(I(t_2)\right), \\ &= u_{ij}\left(I_i(t_2)\right) + \varepsilon_{ij}\left(I_i(t_2)\right) \end{aligned} \tag{14}$$

Note that the utility error is known to a consumer when a purchase decision is made. A consumer purchases the product when its expected utility is higher than not purchasing it.

**5. Empirical Application**

We test our proposed model using data from an incentive compatible stated choice experiment where we control for potential confounds discussed in Section 1 (e.g., endogenous group formation). The task involves search and purchase decisions within the music category for individual songs. The participants in the study were undergraduate students at a large northeastern university. The music category was chosen for several reasons. This is an experiential product category that most consumers are familiar with. Consumers are also familiar with accessing aggregate evaluations for songs (e.g., iTunes ratings for songs).

**5.1. Experimental Design**

Our experiment had two phases: Phase 1 (calibration task) and Phase 2 (incentive compatible choice task). Each phase is described below.

In Phase 1, participants listened to 10 songs of different genres and rated each song on a 0-10 scale. We generated two different lists of 10 songs, and each participant was randomly assigned to one of the two lists. All songs had similar average evaluation on iTunes and the order was randomized across participants. Participants were told to rate each song carefully as their ratings would be used in matching them with other participants and that such matching would be useful in the second phase of the experiment.

In Phase 2, respondents made purchase decisions for several unidentified songs (no artist or genre was specified) without listening to the songs.[3] All songs were worth $1.25 on iTunes. A purchase decision for participants meant deciding between receiving the unidentified song or $1 cash (this cash was in addition to participation fee). Participants' decisions were incentive aligned as they were told that one song out of all unidentified songs would be randomly picked, and they would be compensated with either the actual song (if they had chosen to purchase the song) or $1 additional cash.

Each song was described by the following six attributes. Participants were provided with a summary of aggregate evaluations; (1) the average of the song's rating $(R_j^{0M})$ from iTunes on a 0-10 scale and (2) the standard deviation $(\tau_j^M)$ which captures the population heterogeneity in product evaluations. The average of aggregate evaluations had three levels: low (0.5-3.0), medium (3.0-7.0) and high (7.0-9.5). Note that while each level had a range, a respondent saw a randomly chosen value in the range corresponding to a level. For instance, if the average of aggregate evaluations is low in a particular profile, then we assign the respondent to a value that is randomly chosen in the range (0-3). Likewise, the standard deviation of aggregate evaluations

[3] We did not identify the songs to isolate the impact of aggregate evaluations and preference similarity on consumer search decisions. Future research can look at how product attributes (e.g., artists, song genre) may influence search.

had three levels (0.5-1.5, 1.5-3.5, 3.5-4.5), and the respondent saw an actual value randomly chosen within a range corresponding to a given level.

Participants also had access to social information. Respondents were told that previous participants (classified into two social groups of Undergraduates and MBAs) had listened to the same ten songs as they did in Phase 1 and also all unidentified songs that they would be making purchase decisions for. Respondents were provided with (3) similarity in preference between the participant and undergraduates $(a_j^M)$, (4) between the participant and MBAs $(b_j^M)$, and (5) between undergraduates and MBAs $(c_j^M)$, and they were told that these measures had been computed based on their evaluations of the ten songs in Phase 1. We manipulated all three measures for each song. The absolute similarity in preference (i.e., relevance) between the participant and Undergraduates as well as the participant and MBAs had three levels each (0.1-0.3, 0.3-0.7, 0.7- 0.9). The respondent saw a similarity measure for each source which had an absolute value randomly chosen within a range corresponding to a given level, and a sign randomly chosen to be positive or negative (representing similar or dissimilar preference). The similarity in preference between MBAs and Undergraduates (c) was randomly chosen within a range where the covariance of triadic similarity satisfies regularity conditions (Appendix 1).

Finally, (6) standard deviation of evaluations within the two social groups $(\sigma_j^M)$ captures within-group heterogeneity in evaluations. This was set to be equal between the two groups, and fixed to be one-half of the standard deviation of the aggregate evaluations. We did so primarily to reduce the complexity of the problem for respondents.[4] We generated two orthogonal designs

[4] In a pilot study with a convenience sample of 10 students, we found that it was difficult for them to understand all the information in a profile where we had different standard deviations for each of the two social groups. As our primary goal is to understand how homophily and balance impact consumers' search and purchase behavior, we believe that the lack of orthogonal manipulation of the within-group heterogeneity should have little impact.

of 18 profiles (unidentified songs) from a full factorial design using Proc Optex in SAS. Each participant in the study was assigned to one of the two designs.

Respondents made two decisions for each unidentified song. Based on the aggregate evaluations of the song and preference similarity measures, they decided how many individual ratings to acquire from each social group (search decision). To prevent them from acquiring an arbitrary large number of evaluations, we imposed a cost of search: they had to wait for half a second to retrieve each individual evaluation. Figure 3a shows an example of the search decision interface. After respondents made their search decision (and waited for the designated time), they were provided with the average rating of randomly sampled individuals from each social group with the sample size based on the search decision. Finally, respondents made their purchase decision. Figure 3b shows an example of the purchase decision interface.

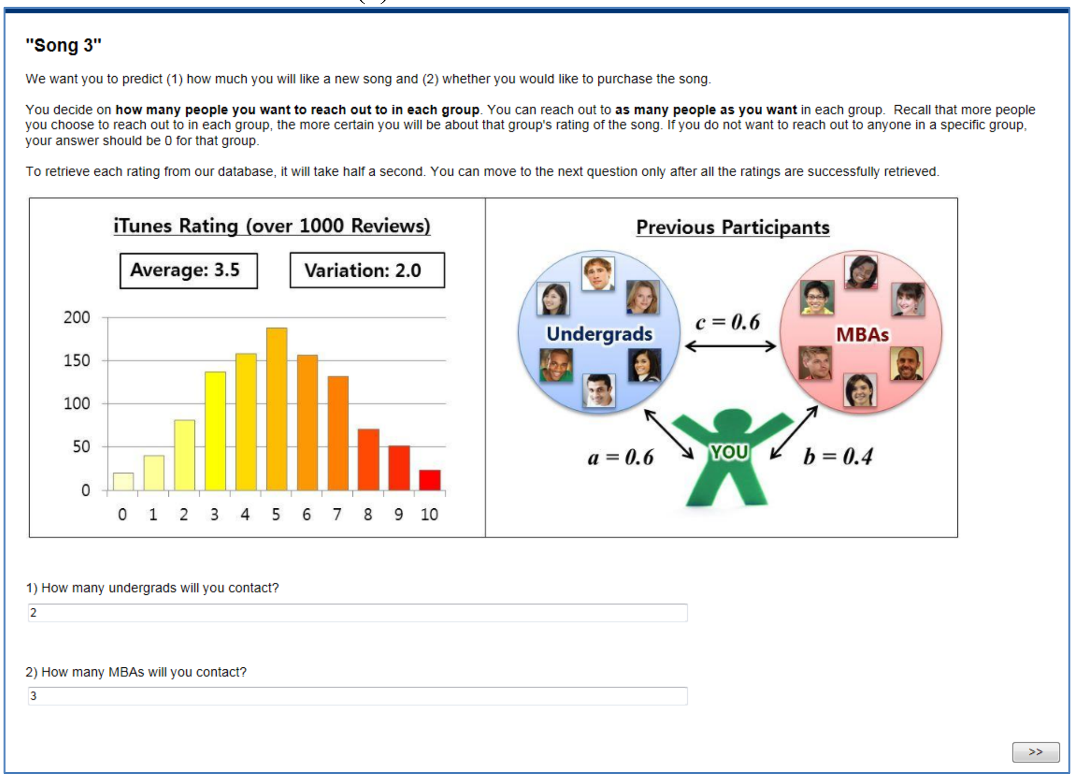
We collected 3,492 (=194 subjects × 18 profiles) pairs of search (how many evaluations to acquire) and purchase decisions (whether or not to purchase a song). Table 1 provides the summary statistics of search and purchase decisions. The average total search amount is around 10 evaluations per song and there appears to be no systematic difference in the search amount across the two groups (as expected since we manipulated the relevance of the two groups across profiles). In 25% of observations, respondents chose to purchase a song.

Table 1. Summary Statistics of Search and Purchase Decision

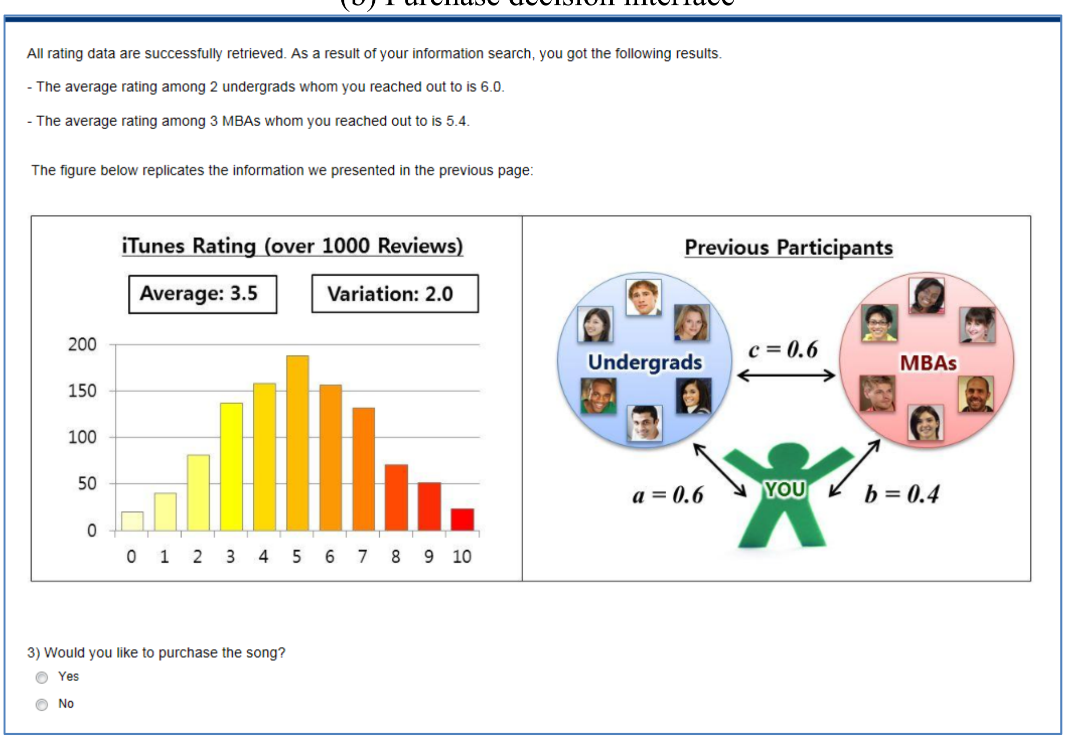
| | Mean | SD | 2.5% | 25% | 50% | 75% | 97.5% |
|---|---|---|---|---|---|---|---|
| Search from Undergrads | 5.2 | 13.1 | 0 | 2 | 4 | 5 | 20 |
| Search from MBAs | 5.1 | 11.8 | 0 | 2 | 4 | 5 | 20 |
| Total Amount of Search | 10.3 | 24.3 | 0 | 5 | 9 | 10 | 40 |
| Purchase Decision (0 is no purchase; 1 is purchase) | 0.25 | 0.43 | | | | | |

Figure 3. Screenshot of Survey Interface

(a) Search decision interface




(b) Purchase decision interface



## 5.2. Empirical Model Specification

We use the experimental data to estimate the model proposed in Section 4. As a baseline model, we fit an SMBL model where people update their belief about own product evaluation according to Bayes rule applied to the *objective* (manipulated) values of all six attributes ( $R_j^{0M}$ , $\tau_j^M$ , $a_j^M$ , $b_j^M$ , $c_j^M$ , and $\sigma_j^M$ ). The SMBL model has two key limitations. First, the informational benefit from positively and negatively relevant sources is forced to be identical (See Appendix 2). Thus, positive feedback from similar others and negative feedback from dissimilar others are equivalent. Second, the informational benefit is always greater under imbalance than balance. Moreover, greater (smaller) redundancy leads to greater informational benefit under imbalance (balance) (See Appendix 2). The intuition behind the latter properties is as follows. Suppose there are two random variables, $Z_1$ and $Z_2$. The variance of $Z_1$+$Z_2$ will be smaller when the two random variables have a negative rather than a positive correlation. Analogously, the greater absolute value of positive (negative) correlation will increase (decrease) the variance of $Z_1$+$Z_2$. This simple intuition applies to SMBL - redundancy cancels the noise of signals under imbalanced condition but amplifies it under a balanced condition, so informational benefit is always greater under imbalance.[5] Obviously, these are restrictive properties, so we cannot either incorporate or test the impact of key features of interpersonal relationships from this model.

The second model we propose and estimate is an extended multivariate Bayesian learning (EMBL) model that incorporates features of social relationships among information recipients and sources. We assume that consumers update their belief about their own evaluation still according to Bayes rule but rely on *subjective* values of attributes. The additional step of

[5] The intuition also applies to the modern portfolio theory (Markowitz 1952). For a given level of return, the overall risk (variance) of a portfolio can be reduced by investing in assets with negative correlation because the poor performance of one asset can be offset with the good performance of another.

allowing for subjective interpretation from consumers provides substantial model flexibility. In particular, the subjective values of attributes depend on the type of relationship that recipients have with their sources - the sign of relevance of each group ($neg_j^A$ and $neg_j^B$, which is 1 when a respondent has negative relevance with Group A or Group B, respectively, and is 0 otherwise) to assess the impact of homophily and the balance status in the triad ($imb_j$, which is 1 if the relationship is imbalanced and is 0 for balance) to accommodate consumers' attitude towards balance. Table 2 summarizes the most flexible specification of EMBL model.[6]

Table 2. Specification of Subjective Attributes

(a) Related to Informational Benefit ($\Theta_i$)

| Subjective Attributes | Specification |
|---|---|
| Aggregate Mean $\left(R_{ij}^0\right)$ | $R_{ij}^0 = R_j^{0M}$ |
| Aggregate Standard Deviation $\left(\tau_{ij}\right)$ | $\tau_{ij} = \tau_j^M \times \exp\left(\theta_{i0}^{pri}\right)$ |
| Within Group Standard Deviation $\left(\sigma_{ij}^A, \sigma_{ij}^B\right)$ | $\sigma_{ij}^A = \sigma_j^M \times \exp\left(\theta_{i0}^{sig} + \theta_{i1}^{sig} neg_j^A + \theta_{i2}^{sig} imb_j\right)$, $\sigma_{ij}^B = \sigma_j^M \times \exp\left(\theta_{i0}^{sig} + \theta_{i1}^{sig} neg_j^B + \theta_{i2}^{sig} imb_j\right)$ |
| Similarity with Groups $\left(a_{ij}, b_{ij}\right)$ | $a_{ij} = a_j^M \times \exp\left(\theta_{i0}^{rel} + \theta_{i1}^{rel} neg_j^A\right)$, $b_{ij} = b_j^M \times \exp\left(\theta_{i0}^{rel} + \theta_{i1}^{rel} neg_j^B\right)$ |
| Similarity between Groups $\left(c_{ij}\right)$ | $c_{ij} = c_j^M \times \exp\left(\theta_{i0}^{red} + \theta_{i1}^{red} imb_j\right)$ |

(b) Related to Cost of Search ($\Delta_i$)

| Subjective Attributes | Specification |
|---|---|
| Cost of Search $\left(k_{ij}^A, k_{ij}^B\right)$ | $k_{ij}^A = \exp\left(\delta_{i0} + \delta_{i1} neg_j^A + \delta_{i2} imb_j\right) \times 10^{-3}$, $k_{ij}^B = \exp\left(\delta_{i0} + \delta_{i1} neg_j^B + \delta_{i2} imb_j\right) \times 10^{-3}$ |

We begin with a description of subjective attributes related to informational benefit from others' evaluations (Table 2a). For consumer *i*, let $\Theta_i$ denote the vector of all related individual-level parameters. Subjective relevance ($a_{ij}$, $b_{ij}$) is specified as a function of the sign of relevance.

[6] In our specification, subjective attribute values are proportional to objective attribute values, and we estimate the proportions. Subjective similarity measures are imposed to have the same sign as objective similarity. We also fit a model where we did not impose such a restriction. All substantive results remained unchanged.

If consumer *i* perceives the information from dissimilar others to be *less* relevant than similar others (e.g., Suls et al. 2000), then $\theta_{i1}^{rel}$ will be significantly negative. Subjective redundancy ($c_{ij}$) is specified as a function of balance status. The term captures consumer *i*'s belief's about the impact of redundancy and balance on the informational benefit. If consumers ignore the interdependency between the two groups (e.g., Hedesstrom et al. 2006), then parameter $\theta_{i0}^{red}$ will take a large negative value (i.e., subjective interdependency will be close to 0). Also, if consumers discount the informational benefit under imbalance by internally reducing the inconsistency in a social system (e.g., Neuberg and Newsom 1993), parameter $\theta_{i1}^{red}$ will be significantly negative.

Subjective signal variance $\left(\sigma_{ij}^{A}, \sigma_{ij}^{B}\right)$ is specified as a function of both sign of relevance and balance status. For instance, if people believe that the information from dissimilar others is less diagnostic (i.e., has higher variance), the parameter $\theta_{i1}^{sig}$ will be significantly positive. If people subjectively perceive inconsistent information as less diagnostic (e.g., Tversky and Kahneman 1974), then the parameter $\theta_{i2}^{sig}$ will be significantly positive.

Unlike the attributes related to evaluations from social sources, the subjective aggregate distribution of evaluations about a specific song ( $R_{ij}^{0}$ and $\tau_{ij}$ ) is unlikely to be more or less dispersed depending on their similarity of preference with either social group. Therefore, we do not specify the subjective aggregate distribution as functions of relevance or balance status. Additionally, we cannot identify both subjective mean and subjective variance of aggregate distribution jointly (See Appendix 4), so we assume the subjective average is the same as the manipulated average while the subjective standard deviation is proportional to the manipulated standard deviation. We empirically validate this specification in Appendix 3.

In Table 2b, we summarize the specification of subjective attributes related to the cost of search. For consumer *i*, let $\Delta_i$ denote a vector of all related individual-level parameters. We manipulated the cost of search as the time that respondents have to wait to acquire a single evaluation. Therefore, the parameter $\delta_{i0}$ captures consumer *i*'s (baseline) unit cost of search (wait time) on the utility scale. If the consumer has greater cost of search from dissimilar others (Carley 1991), the parameter $\delta_{i1}$ will be significantly positive. If a consumer has either greater discomfort (Heider 1946) or greater cost for information processing (Meyers-Levy and Tybout 1989) under an imbalanced social system, the parameter $\delta_{i2}$ will be significantly positive.

Finally, as other product related attributes were not included and the price of song was held fixed, the parameter $\alpha_{ij}$ in Equation 1 is the parameter $\gamma_{i0}$ in our empirical model. Let $\Gamma_i$ denote the vector of utility parameters $\left(\gamma_{i0}, \ln(\beta_i)\right)$.

### 5.3. Estimation

We have 194 respondents (*i* =1…*N*), and each made search and purchase decisions for 18 songs (*j* =1…*J*). For respondent *i* and song *j*, let $n_{ij}^*$ denote the actual search decision and let $y_{ij}$ be an indicator variable that takes a value of 1 if he decides to purchase the song and is 0 otherwise.

We make the following distributional assumptions on the two errors in our model. First, the search utility error ( $\xi_{n_{ij}}$ ) follows IID Type I Extreme value distribution with a scale parameter $\lambda_i$. Second, purchase utility error follows a standard Normal distribution. Then, the conditional likelihood that a consumer *i* makes a search decision of $n_{ij}^*$ and purchase decision of $y_{ij}$ for a song *j* can be expressed as:

$$\Pr\left(n_{ij}^*, y_{ij} \mid \Gamma_i, \Delta_i, \Theta_i, \lambda_i\right) = \Pr\left(n_{ij}^* \mid \Gamma_i, \Delta_i, \Theta_i, \lambda_i\right) \times \Pr\left(y_{ij} \mid n_{ij}^*, \Gamma_i, \Theta_i\right), \qquad (15)$$

where the first term on the right-hand side is the search likelihood which follows multinomial logit and the second term is the purchase likelihood which follows binary Probit.

In the experiment, we did not constrain the search amount and any combination of two non-negative integers is a possible option. Clearly, it is not feasible to estimate the model as is. We use a subset of options to estimate the model by the positive conditioning property (McFadden 1978; Train et al. 1987). For consumer *i* and song *j*, let $W_{ij}$ denote a consideration set that includes the actual search option ($n_{ij}^*$) and 5 other possible options of $n_{ij}$, which are randomly selected from the empirical distribution of search decisions in our dataset.[7] Then, we can write the search likelihood as:

$$\Pr\left(n_{ij}^* \mid \Gamma_i, \Delta_i, \Theta_i, \lambda_i\right) = \frac{\exp\left(v_{ij}^b\left(n_{ij}^* \mid I_i(t_1), \Gamma_i, \Theta_i\right)\big/\lambda_i - v_{ij}^c\left(n_{ij}^* \mid \Delta_i\right)\big/\lambda_i\right) \times h\left(W_{ij} \mid n_{ij}^*\right)}{\sum_{n_{ij} \in W_{ij}} \exp\left(v_{ij}^b\left(n_{ij} \mid I_i(t_1), \Gamma_i, \Theta_i\right)\big/\lambda_i - v_{ij}^c\left(n_{ij} \mid \Delta_i\right)\big/\lambda_i\right) \times h\left(W_{ij} \mid n_{ij}\right)}, \tag{16}$$

where $h\left(W_{ij} \mid n_{ij}\right)$ denotes a bias adjustment factor to account for using a subset of options. Specifically, $h\left(W_{ij} \mid n_{ij}\right)$ is the probability that consumer *i* formed a consideration set of $W_{ij}$ given that he made a search decision of $n_{ij}$. We used importance sampling and computed the bias adjustment factors from the empirical distribution of search decisions. Lastly, note that $v_{ij}^b(\cdot)$ does not have a closed form expression (see Section 4.2.3) and is computed using a Monte-Carlo simulation.

The conditional purchase likelihood is a binary Probit likelihood specified as:

$$\Pr\left(y_{ij} \mid n_{ij}^*, \Gamma_i, \Theta_i\right) = \Phi\left(u_{ij}\left(I_i(t_2) \mid \Gamma_i, \Theta_i\right)\right)^{y_{ij}} \left(1 - \Phi\left(u_{ij}\left(I_i(t_2) \mid \Gamma_i, \Theta_i\right)\right)\right)^{1-y_{ij}}, \tag{17}$$

[7] We also estimated a model where $W_{ij}$ consists of 10 alternatives including the observed search decision. All substantive findings remained unchanged.

where the information set in the second stage $\left(I_i(t_2)\right)$ includes the search decision made in the first stage $\left(n_{ij}^*\right)$.

Given that there are common parameters $\left(\Gamma_i, \Theta_i\right)$ in both stages, the two decisions are estimated jointly. Therefore, the conditional likelihood of observing the decisions for consumer *i* for all *J* songs is:

$$L_i \mid (\Gamma_i, \Delta_i, \Theta_i, \lambda_i) = \prod_{j=1...J} \Phi\left(u_{ij}\left(I_i(t_2) \mid \Gamma_i, \Theta_i\right)\right)^{y_{ij}} \left(1-\Phi\left(u_{ij}\left(I_i(t_2) \mid \Gamma_i, \Theta_i\right)\right)\right)^{1-y_{ij}} \\ \times \frac{\exp\left(v_{ij}^b\left(n_{ij}^* \mid I_i(t_1), \Gamma_i, \Theta_i\right)\big/\lambda_i - v_{ij}^c\left(n_{ij}^* \mid \Delta_i\right)\big/\lambda_i\right) \times h\left(W_{ij} \mid n_{ij}^*\right)}{\sum_{n_{ij} \in W_{ij}} \exp\left(v_{ij}^b\left(n_{ij} \mid I_i(t_1), \Gamma_i, \Theta_i\right)\big/\lambda_i - v_{ij}^c\left(n_{ij} \mid \Delta_i\right)\big/\lambda_i\right) \times h\left(W_{ij} \mid n_{ij}\right)}. \tag{18}$$

To capture consumer heterogeneity, individual-level parameters ($\Gamma_i, \Delta_i, \Theta_i, \ln(\lambda_i)$) are assumed to be distributed multivariate normal with mean vector $\left(\Gamma, \Delta, \Theta, \ln(\lambda)\right)$ and covariance matrix Σ. The unconditional likelihood *L* for a sample of *N* customers is:

$$L = \prod_{i=1...N} \int L_i \mid (\Gamma_i, \Delta_i, \Theta_i, \lambda_i)\; dF\left(\Gamma_i, \Delta_i, \Theta_i, \lambda_i \mid \Gamma, \Delta, \Theta, \lambda, \Sigma\right). \tag{19}$$

where $F\left(\Gamma_i, \Delta_i, \Theta_i, \lambda_i \mid \Gamma, \Delta, \Theta, \lambda, \Sigma\right)$ denotes the multivariate normal density function.

The model parameters are estimated using standard hierarchical Bayesian Markov chain Monte Carlo (MCMC) methods. We use the following set of priors for all population level parameters. Let $\left(\Gamma, \Delta, \Theta, \ln(\lambda)\right)$ be a $p \times 1$ vector and that $\Sigma^{-1}$ is a $p \times p$ matrix. Then, the prior for $\left(\Gamma, \Delta, \Theta, \ln(\lambda)\right)$ is a multivariate normal with mean of 0 and covariance of $p \times p$ identity matrix. The prior for Σ is a Wishart distribution where the scale matrix is $p \times p$ identity matrix, and $p+1$ degrees of freedom. The details of the full conditional distributions are available from the authors upon request. Appendix 4 outlines the identification of parameters.

## 6. Results

### 6.1. Model Performance Results

In Section 5.2, we outlined a model specification of a flexible EMBL model where multiple features of interpersonal relationships were incorporated. For instance, homophily could affect search and purchase decisions by changing the credibility of signals, relevance of signals, or cost of search. A flexible model may, however, suffer from over-specification.

To determine the final model specification, we fit 9 model variants, which differed on how interpersonal aspects are incorporated in the model. Table 3 summarizes the specifications of all models. In Table 3, "M" denotes that the update is based on the objective row attribute as we "manipulated", "S" denotes that the update is based on the "subjective" row attribute which is not associated with either the sign of relevance or balance status, and "F" denotes that the update is based on subjective row attribute which is a "function" of the sign of relevance and balance.

The most restrictive model was Model 1 where consumer learning was based on objective attributes as we manipulated (SMBL). This was the baseline model. The most general model (Model 9) followed the specification we outlined in Section 5.2. Other models (Models 2-8) differed on how interpersonal aspects are incorporated in various components.

In Model 2, the update was based on four subjective attributes (i.e., subjective prior variance, subjective signal variance, subjective similarity structure, and subjective cost of search), but none of them was a function of either the sign of relevance $\left(neg_j^A, neg_j^B\right)$ or the balance status $\left(imb_j\right)$. In Models 3-5, all four attributes (prior variance, signal variance, similarity structure, and cost of search) are subjective, and in each model, we made one of the attributes a function of the sign of relevance and balance status. To be specific, we allowed signal variance (Model 3), similarity structure (Model 4) and cost of search (Model 5) to depend on the sign of relevance and balance status. In Models 6-8, all four attributes remained subjective, and we made two of four subjective attributes a function of the sign of relevance and balance status. Specifically, we allowed signal variance and similarity structure (Model 6), signal variance and cost of search (Model 7) and similarity structure and cost of search (Model 8) as functions of relevance and balance status.

Table 3. Model Comparisons

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 |
|---|---|---|---|---|---|---|---|---|---|
| (a) Specification | | | | | | | | | |
| Prior Mean | M | M | M | M | M | M | M | M | M |
| Prior Variance | M | S | S | S | S | S | S | S | S |
| Signal Variance | M | S | F | S | S | F | F | S | F |
| Similarity Structure | M | S | S | F | S | F | S | F | F |
| Cost of Search | S | S | S | S | F | S | F | F | F |
| (b) Within-Sample Fit | | | | | | | | | |
| DIC | 14,600 | 13,738 | 13,446 | 13,640 | 12,949 | 13,257 | 12,912 | 12,857 | 12,855 |
| BPIC | 14,997 | 14,236 | 13,983 | 14,114 | 13,563 | 13,874 | 13,539 | 13,487 | 13,503 |
| WAIC | 14,688 | 13,988 | 13,898 | 13,984 | 13,327 | 13,579 | 13,460 | 13,269 | 13,369 |
| Purchase Hit Rate | 0.83 | 0.84 | 0.84 | 0.84 | 0.82 | 0.81 | 0.82 | 0.81 | 0.80 |
| Search Hit Rate (9 groups) | 0.37 | 0.40 | 0.41 | 0.40 | 0.43 | 0.43 | 0.43 | 0.44 | 0.45 |
| Search Hit Rate (25 groups) | 0.26 | 0.28 | 0.29 | 0.28 | 0.30 | 0.31 | 0.30 | 0.32 | 0.32 |
| Search MED | 4.39 | 3.98 | 3.73 | 3.99 | 3.10 | 3.30 | 3.12 | 2.95 | 3.01 |
| (c) Holdout-Sample Fit | | | | | | | | | |
| Validation LL | -1,521 | -1,552 | -1,529 | -1,576 | -1,465 | -1,462 | -1,465 | -1,447 | -1,446 |
| Purchase Hit Rate | 0.79 | 0.80 | 0.81 | 0.81 | 0.76 | 0.76 | 0.76 | 0.76 | 0.76 |
| Search Hit Rate (9 groups) | 0.37 | 0.37 | 0.39 | 0.38 | 0.41 | 0.42 | 0.41 | 0.42 | 0.41 |
| Search Hit Rate (25 groups) | 0.25 | 0.24 | 0.26 | 0.26 | 0.26 | 0.27 | 0.28 | 0.28 | 0.27 |
| Search MED | 4.38 | 4.07 | 3.78 | 4.06 | 3.22 | 3.49 | 3.29 | 2.98 | 3.18 |

*Note: In Table 3, "M" denotes that the update is based on the objective row attribute as we "manipulated", "S" denotes that the update is based on the "subjective" row attribute which is not associated with either the sign of relevance or balance status, and "F" denotes that the update is based on subjective row attribute which is a "function" of the sign of relevance and balance status.

We used MCMC methods for estimating all 9 model variants. For each model, we ran sampling chains for 200,000 iterations. In each case, convergence was assessed by monitoring the time series of the draws. We report the results based on 100,000 draws retained after discarding the initial 100,000 draws as burn-in iterations. For each participant, we randomly select 15 of the 18 song profiles for model estimation and use the remaining 3 for out-of-sample prediction.

We compared the models on multiple measures of model fit. First, we report Bayesian model selection criteria such as DIC (Spiegelhalter et al. 2002), BPIC (Ando 2007), and WAIC (Watanabe 2010). These measures are used to evaluate within-sample fit and complexity of each model. Smaller numbers denote a better model. Second, we computed multiple measures of within-sample model prediction. To assess the prediction for purchase decision, we used purchase hit-rate where cut-off was fixed at 0.5. To evaluate the model prediction for search decision, we used hit rate and median Euclidean distance (MED).[8] Finally, we computed validation log-likelihood (VLL) in the holdout sample to assess predictive validity (Montoya, Netzer, and Jedidi 2010; Iyengar and Jedidi 2012).

A comparison of models on the various criteria shows that Model 8 generally outperforms the baseline model (SMBL) as well as other EMBL model variants. Specifically, in comparison to the most flexible model (Model 9), Model 8 is superior on nine out of twelve criteria. On the remaining three criteria, the differences between the two models are small. Thus, on the basis of parsimony, we use Model 8 as our main model for the rest of our discussion.

### 6.2. Parameter Estimates

As is common in Bayesian analysis, we summarize the posterior distribution of the parameters by reporting their posterior means and 95% posterior confidence intervals. Table 4 contains the

[8] For computing the search hit rate, we *discretized* the observed $n_{ij}^{A^*}$ and $n_{ij}^{B^*}$ into 3 levels (5 levels) each based on their quartiles and thus the overall search decision, which is a combination of $n_{ij}^{A^*}$ and $n_{ij}^{B^*}$, is classified into 9 options (25 options). The search hit rate is the proportion of observations where the observed search option matches the option with the highest search utility based on our model estimates. The MED measure compares the model prediction for the search decision with the actual search decision on a *continuous* scale. We computed the Euclidean distance between actual and predicted search decision for each observation, and report the median among all the observations.

model estimates from the proposed model (Model 8) and from the baseline model (Model 1). All other model estimates are available from the authors upon request.

Table 4. Model Estimates for Baseline Model and Proposed Model

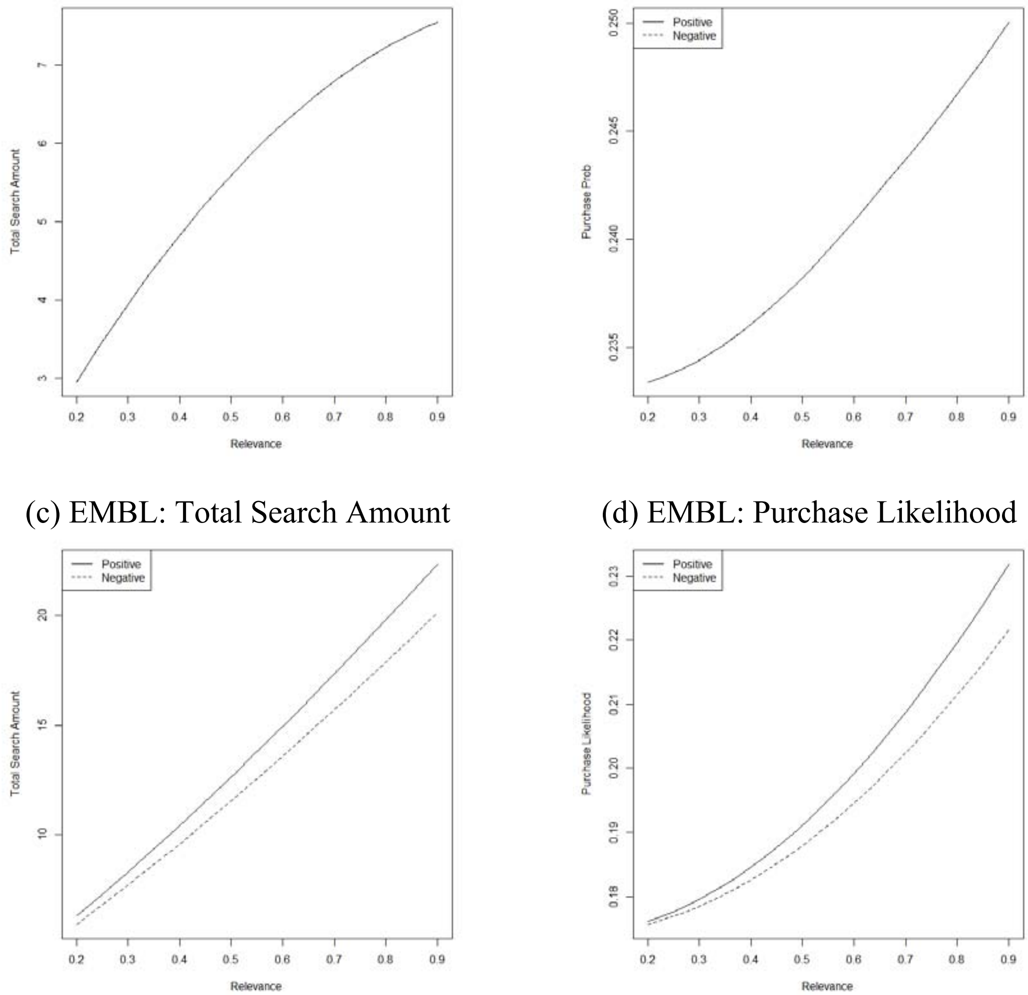
| Population Parameter Estimates | Baseline Model (SMBL) | Proposed Model (EMBL) |
|---|---|---|
| Utility Parameters: Intercept ($\gamma_0$) | -0.06<br>(-0.26, 0.16) | **1.37***<br>( 1.15, 1.63) |
| Utility Parameters: Rating ($Ln(\beta)$) | **1.01***<br>( 0.82, 1.22) | **0.69***<br>( 0.31, 1.09) |
| Relevance: Base ($\theta_0^{rel}$) | | **-0.27***<br>(-0.40, -0.17) |
| Relevance: Negative ($\theta_0^{rel}$) | | **-0.19***<br>(-0.35, -0.03) |
| Redundancy: Base ($\theta_0^{red}$) | | **-0.90***<br>(-1.32, -0.39) |
| Redundancy: Imbalance ($\theta_1^{red}$) | | 0.08<br>(-0.62, 0.53) |
| Prior Standard Deviation: Base ($\theta_0^{pri}$) | | 0.33<br>(-0.06, 0.70) |
| Signal Standard Deviation: Base ($\theta_0^{sig}$) | | **0.92***<br>( 0.49, 1.36) |
| Cost: Base ($\delta_0$) | **1.01***<br>( 0.53, 1.49) | **0.72***<br>( 0.19, 1.21) |
| Cost: Negative ($\delta_1$) | | 0.24<br>(-0.09, 0.54) |
| Cost: Imbalance ($\delta_2$) | | **0.68***<br>( 0.40, 0.95) |
| Scale Parameter (Ln($\lambda$)) | **-3.38***<br>(-3.89, -2.87) | **-3.40***<br>(-3.89, -2.87) |

Note: * denotes significance. 95% confidence intervals are errors in parentheses.

Our main model provides interesting insights. First, people perceive signals from negatively relevant source (i.e., dissimilar others) to be significantly less informative than a positively relevant source: Positive relevance is discounted by 24% (=1-exp(-0.27)) and negative relevance is discounted by 37% (=1-exp(-0.27-0.19)). In contrast, the sign of relevance does not have a significant effect on perceived cost (mental cost) of information seeking. The finding is consistent with the past studies on homophily, and further pinpoints the driver behind the observed phenomenon; homophily in social learning is driven by informational benefit (Suls et al. 2000) rather than cost of information seeking.

Figure 4. Simulated Purchase and Search Decisions: Positive and Negative Relevance

(a) SMBL: Total Search Amount (b) SMBL: Purchase Likelihood




Note: For generating the figures, we only varied the relevance of one source $\left(a_j^M\right)$ and fixed all other attributes ( $R_j^{0M} = 5$ , $\tau_j^M = 5$ , $\sigma_j^M = 2.5$ , $b_j^M = 0.2$ , and $c_j^M = 0.0$ ). The figures show the search amount and purchase likelihood averaged across all respondents.

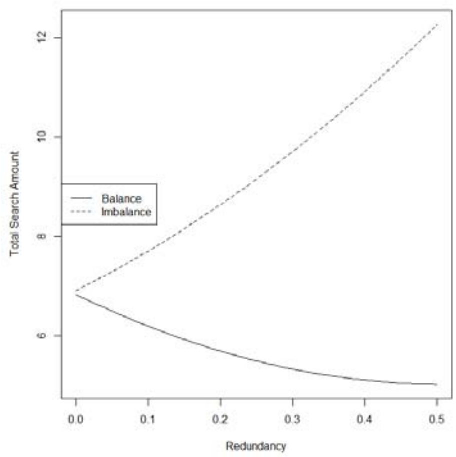
In Figure 4, we illustrate how the sign of relevance (with all other attributes fixed) impacts consumers' search and purchase decisions. For these figures, given the attribute values and the estimates of individual parameters, we simulated the purchase likelihood and total search amount for each respondent under both SMBL and EMBL models. The plots show the averages across all respondents. In the SMBL specification, where the informational benefit from similar and dissimilar others is forced to be identical, the sign of relevance does not have an impact on either search or purchase decision (Figure 4a and 4b). In the proposed EMBL specification, where the informational benefit is greater from similar others than dissimilar others, simulation results show that people tend to search more from similar others than dissimilar others (Figure 4c), and purchase more when the information is collected from similar others than dissimilar others (Figure 4d).

Second, imbalance has a nuanced effect on social learning. On the one hand, the status of balance does not have a significant effect on perceived redundancy. The finding implies that people do not discount the informational benefit under imbalanced system. That is, consumers find greater informational benefit under imbalance than balance (Goethals and Nelson 1973; Levine, Resnick, and Higgins 1993). On the other hand, people have significantly greater subjective cost of gathering information under imbalance than balance – the subjective cost of search is almost double (= exp(0.68+0.72)/exp(0.72)) under imbalance as compared to balance. The finding may be driven by either discomfort with an imbalanced relationship (Heider 1946) or greater cost of information processing (Meyers-Levy and Tybout 1989).

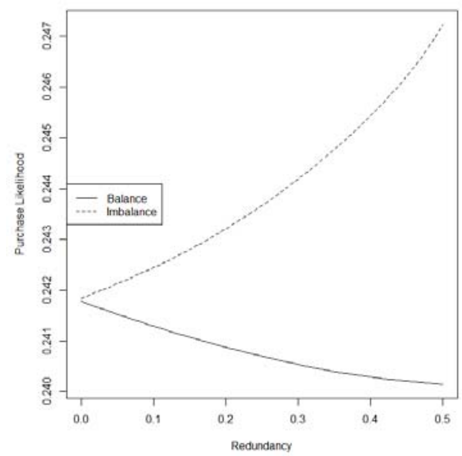
In Figure 5, we show how triadic balance (given all the other attributes fixed) impacts search and purchase decisions. As before, given the attribute values and the estimates of individual parameters, we simulate the purchase likelihood and total search amount for each respondent under both model specifications. The plots show the averages across all respondents. In the SMBL specification, where the informational benefit is forced to be greater under imbalance, the search amount and the purchase likelihood is always greater under imbalance (Figure 5a and 5b). In the proposed EMBL model, however, people search less under imbalance because of greater subjective cost of search (Figure 5c). Interestingly, the lower amount of search under imbalance still leads to greater purchase likelihood (Figure 5d) as compared to the balanced condition because people can reduce their uncertainty to a greater extent under the former. In sum, the impact of imbalance on the search decision is dominated by cost effect, but the impact on the purchase decision is dominated by informational benefit.

Figure 5. Simulated Purchase and Search Decisions: Balanced and Imbalanced Relationships

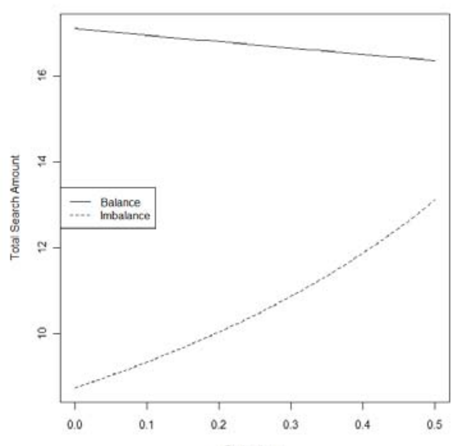
(a) SMBL: Total Search Amount



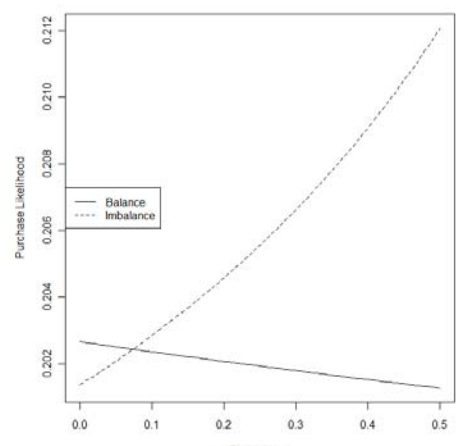

(b) SMBL: Purchase Likelihood




(c) Main(EMBL): Total Search Amount




(d) Main(EMBL): Purchase Likelihood




Note: For generating the figure, we only varied the relevance of one source $\left(c_j^M\right)$ and fixed all other attributes ($R_j^{0M} = 5$, $\tau_j^M = 5$, $\sigma_j^M = 2.5$, $a_j^M = 0.6$, and $b_j^M = 0.3$). The figures show the search amount and purchase likelihood averaged across all respondents.

## 7. Managerial Implications

Many Internet retailers such as goodreads.com or Netflix provide a fixed number of friends' reviews to help consumers reduce their uncertainty about a product or service. For instance, Netflix provides nine friends' reviews as default. Similarly, Foursquare collects a commission from restaurants when consumers, after reading their friends' reviews, reserve via their portal. The hope is that any reduction in uncertainty will translate to an increase in sales. In this section, we show how firms should provide social reviews to achieve their goal.

We make several assumptions. First, both the firm and consumers know that there are two groups of friends (Group A and B) and which friend is in what group. Moreover, both know the similarity structure in a social system – a firm knows the similarity of preference from past data of product evaluations, and a consumer knows it from past experience. Second, a consumer may or may not read all the reviews on the screen: if the number of reviews exceeds his need given the cost of search, he will not read all reviews. Third, for a product $j$, there is a marginal cost ($cost_j$) to the firm for providing each additional review. The cost captures the fact that an inclusion of too many reviews may harm the look of a website, or reduce the space for other needs.

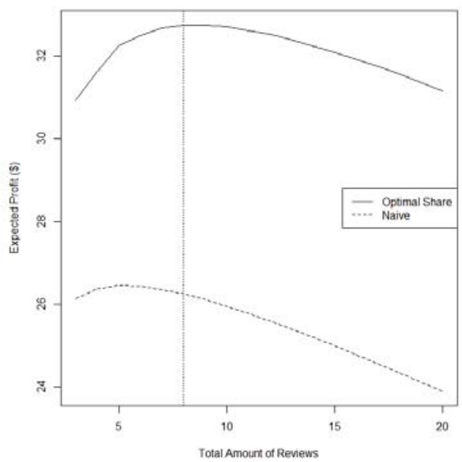
Given that the firm provides social information to increase expected profit from a product $j$, the firm's objective function can be written as:

$$\begin{aligned}\left(total_j^*, share_j^{A*}\right) &= \arg\max_{\left(total_j, share_j^A\right)} E\left[\sum_{i=1}^{N} price_j \times y_{ij}\left(total_j, share_j^A\right)\right] - cost_j \times total_j, \quad (20)\\ &= \arg\max_{\left(total_j, share_j^A\right)} \sum_{i=1}^{N} price_j \times \Pr\left(y_{ij} = 1 \mid total_j, share_j^A\right) - cost_j \times total_j,\end{aligned}$$

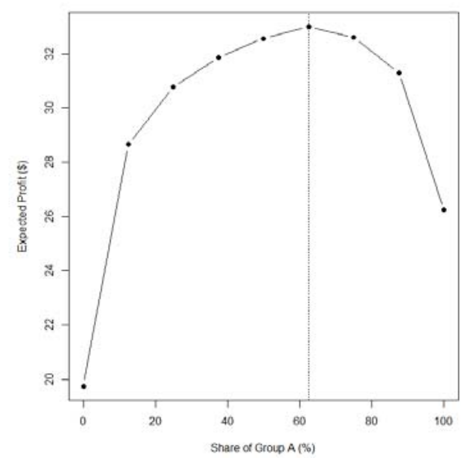
where $i$ denotes a consumer ($i$=$1 \ldots N$), $price_j$ denotes the price of the product $j$, $total_j$ denotes the total number of reviews to put on the screen ($total_j = n_j^A + n_j^B$), $share_j^A$ denotes the share of reviews from Group A $(share_j^A = n_j^A / total_j)$, and $y_{ij}$ is an indicator of purchase that takes a value of 1 when a consumer *i* purchases a product *j* and is 0 otherwise. Additionally, the marginal cost for the product is assumed to be zero.

Figure 6. Optimal Policy: Positive vs Negative Relevance

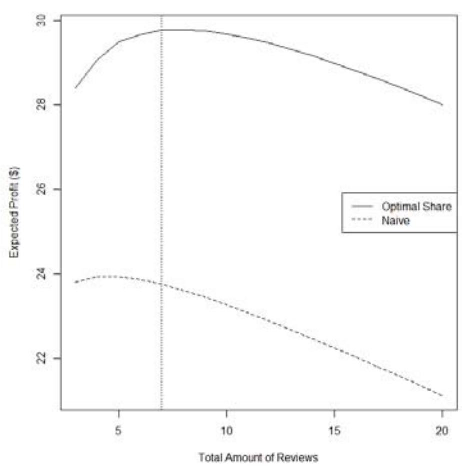
(a) Positive Relevance: Optimal Number of Reviews



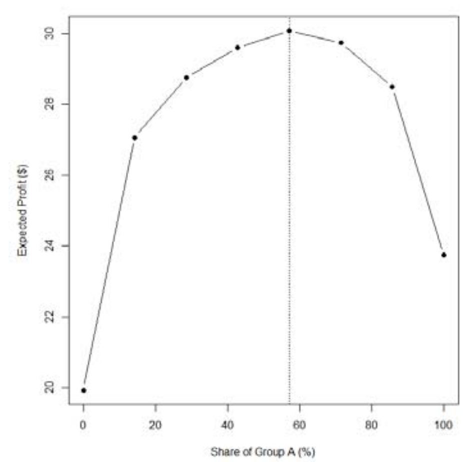

(b) Positive Relevance: Optimal Share




(c) Negative Relevance: Optimal Number of Reviews




(d) Negative Relevance: Optimal Share




Note: Figure 6a and 6c show the relationship between the total number of reviews and the expected profit when a firm provides the optimal share of reviews given the total number of reviews. Figure 6b and 6d show the relationship between the share of reviews and the expected profit when a firm provides the optimal number of reviews. Every attribute related to consumer learning is fixed to be the same between Figure 6a, 6b and 6c, 6d ($R_j^{0M} = 5$, $\tau_j^M = 5$, $\sigma_j^M = 2.5$, $b_j^M = 0.3$, and $c_j^M = 0.0$), but for the sign of relevance with group A. ($a_j^M = 0.7$ in Figure 6a, 6b and $a_j^M = -0.7$ in Figure 6c, 6d).

Using policy experiments, we discuss how firms should provide social information $(total_j^*, share_j^{A*})$ to maximize expected profit. Throughout these exercises, we assume that there are 194 heterogeneous consumers in the market ($N$=194), and their individual parameters are represented by our individual model estimates. The price of the product is fixed at \$1.25 ($price_j$ = \$1.25), and the marginal cost of providing a friend's review is assumed to be 20% of the price ($cost_j$ = \$0.25).[9]

In Figure 6, we show how the sign of relevance affects the optimal number of reviews $(total_j^*)$ and the optimal share $(share_j^{A*})$, and the subsequent expected profits based on our model estimates. To do so, we consider two variants of the social triad where every attribute related to consumer learning is fixed to be the same, but for the sign of relevance with group A (without loss of generality group A is assumed to be the more relevant source).

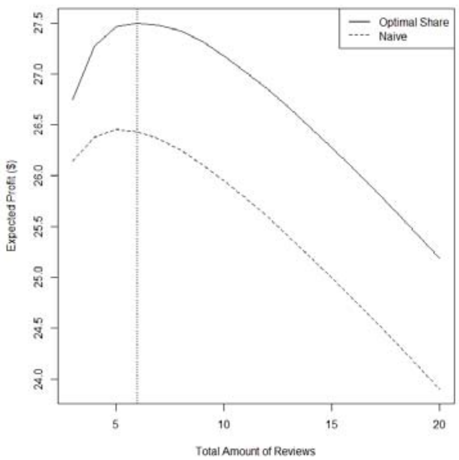
Figures 6a and 6c suggest that the firm should provide more reviews when group A has positive relevance (8 reviews) than negative relevance (7 reviews), and the subsequent expected profit is around 10% (=\$33/30) greater. This result is directly related to our finding of homophily being driven by informational benefit: information from similar others lead to less consumer uncertainty, and to higher expected profit. For the same reason, the optimal share for group A is greater when the relevance is positive (63% in Figure 6b) than negative (57% in Figure 6d). Note that results also indicate that there is a substantial increase in expected profit of around 25% (=\$30/24 or \$33/26) when the firm provides the optimal share of reviews instead of taking a naïve strategy of providing all 10 reviews from the more relevant group (See Figure 6a and 6c). The result reflects that the informational benefit from each group exhibits diminishing marginal returns.

[9] We also performed policy experiments with other costs and all substantive results remained unchanged.

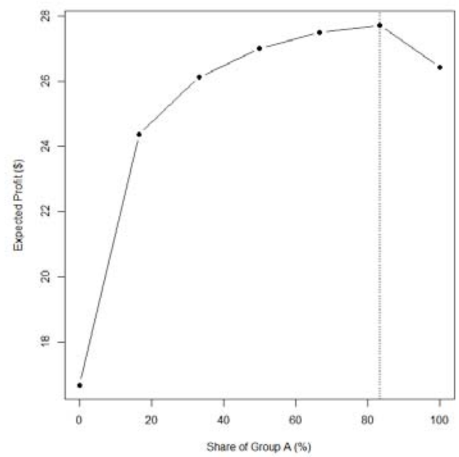
In Figure 7, we determine the impact of balance status on the optimal share and the expected profit. Again, we consider two variants of the social triad where, every attribute is fixed to be the same, but for the sign of redundancy among the two sources. Thus, the status of balance in the social triad differs between the two scenarios.

Figure 7. Optimal Policy: Balance vs Imbalance

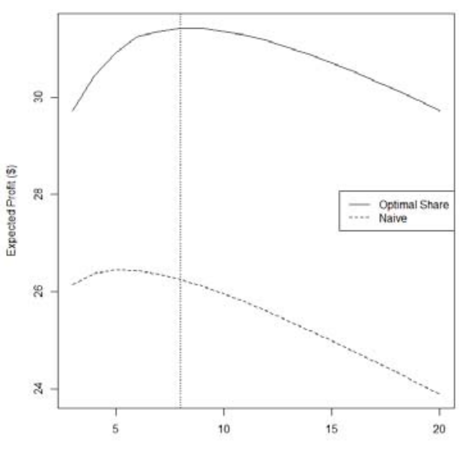
(a) Balance: Optimal Number of Reviews



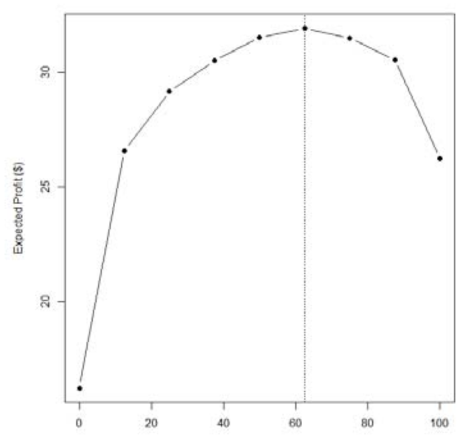

(b) Balance: Optimal Share




(c) Imbalance: Optimal Number of Reviews




(d) Imbalance: Optimal Share




Note: Note: Figure 7a and 7c show the relationship between the total number of reviews and the expected profit when a firm provides the optimal share of reviews given the total number of reviews. Figure 7b and 7d show the relationship between the share of reviews and the expected profit when a firm provides the optimal number of reviews. Every attribute related to consumer learning is fixed to be the same between Figure 7a and 7b ( $R_j^{0M} = 5$ , $\tau_j^M = 5$ , $\sigma_j^M = 2.5$ , $a_j^M = 0.7$ , and $b_j^M = 0.3$ ), but for the sign of redundancy. ( $c_j^M = 0.3$ in Figure 7a and $c_j^M = -0.3$ in Figure 7b). Thus, the status of balance in the social triad differs between the two scenarios.

From Figures 7a and 7c, we see that the firm should provide more reviews under imbalance (8 reviews) than balance (6 reviews), and the subsequent expected profit is around 15% (=$32/28) higher. This finding is related to greater informational benefit under imbalance. It is notable that while consumers would like to search more under balance (see Section 6), the firms should still provide fewer reviews. This is because while an additional review under a balanced system may be easier for consumers to read, it does not sufficiently increase the likelihood of purchase. Note that we assume that the firm has the same cost of providing an additional review regardless of balance.

As shown in Figure 7b and 7d, the optimal share under balance is 83% whereas that under imbalance is 63%. Why do we see this difference? Intuitively, suppose that there are two groups of people who have preferences that are similar to a focal consumer. If preferences of the two groups are similar (a balanced relationship), consumers will expect to receive redundant information from both groups which will add little informational benefit: Providing more reviews from more relevant source can lead to less consumer uncertainty, and therefore to higher expected product profit. In this case, naïve strategy may be a reasonable approximation for optimal strategy: firms can increase the expected profit no more than 4% (=$27/26) by taking optimal strategy (Figure 7a). If preferences of the two groups are dissimilar (an imbalanced relationship), consumers will expect to receive conflicting views from each group. In this case, consumers can compare different viewpoints by reading product reviews from both groups and better determine how much they will like the product. Therefore, diversifying the number of reviews can lead to less uncertainty, and higher expected profit: Firms can increase the expected profit by 20% (=$31/26) by taking optimal strategy as opposed to naïve strategy (Figure 7c).

## 8. Conclusions

Consumers often turn to friends or other social contacts before making product purchases. In this paper, we investigate consumer search and learning from social contacts for the purchase of experiential products. There are two important aspects of our study. First, the learning model is purposive and accommodates information search from consumers for a planned product purchase. Consumers may gather information from their social contacts, who may have product preferences that are systematically different from their own. While the model is grounded in widely-accepted Bayesian learning mechanism, it flexibly accommodates features of interpersonal relationships (e.g., homophily) that impact social learning. Second, search and purchase decisions are modeled as inter-related, but temporally separated, decisions thus allowing us to assess the impact of social relationships on each decision. The model is estimated using incentive compatible stated choice experiments where consumers purchase decisions for individual music tracks while having access to others' evaluations.

Our results provide insights into how consumers learn from others. First, there is clear evidence of homophily in social learning as consumers prefer to gather information from similar others. Our modeling framework allows us to disentangle whether the impact of homophily is due to either the informational benefit of gathering evaluations from similar others or a reduction in the cost of seeking such information. The results suggest that the main driver is the former - consumers find the informational benefit by gathering the reviews from similar others to be greater than those from dissimilar others (Suls et al. 2000). This finding is in contrast to the assumption in standard Bayesian learning models where the informational benefit from similar and dissimilar others are forced to be identical.

Second, the impact of balance status on social learning is more nuanced; people prefer imbalance from a perspective of informational benefit, but balance from a perspective of cost. To be more specific, people understand that informational benefit is greater under an imbalanced social system (Goethels and Nelson 1973; Levine et al. 1993). At the same time, the cost of information-seeking is higher under an imbalanced relationship; it may reflect discomfort under imbalance (Heider 1946) or additional burden for processing incongruent information (Meyers-Levy and Tybout 1989). Our results indicate that both effects are present. People tend to search less under an imbalanced system as compared to a balanced system due to search costs. However, the lower amount of search under imbalance still leads to greater informational benefit as compared to the balanced condition. Note that as the similarity of preference among consumers was manipulated, our evidence of the impact of homophily and balance on social learning does not suffer from confounds such as interpersonal affect or frequency of interactions typically present in observational data.

For marketing practice, our findings yield actionable insights. Recently, several firms (e.g., Opentable) have begun providing social information to help their customers reduce product uncertainty. The top-line message of our study to these firms is that social reviews offered in an imbalanced social system have significant informational benefit. Our policy experiments show that firms can increase their expected profit by providing customers with product reviews from others with whom they have an imbalanced relationship. The benefit of imbalance will be even greater when firms diversify the share of offered reviews from each social group that their customers have. It is also notable that firms should provide fewer reviews in a balanced social system: consumers may want to read more reviews due to lower search cost, but additional reviews (with low informational benefit in a balanced relationship) only marginally increase product sales. Thus, it is critical for firms to understand the relationship among their customers and their social contacts.

There are several promising avenues for future research. First, we can consider how different objectives for social information-seeking can change the search behavior. People often turn to others to form a consideration set (e.g., narrow the set of acceptable movies), and it will be interesting to quantify the differential roles of homophily and balance in consideration set formation. Second, consumers perform different types of information search. We considered simultaneous search in this paper. In other scenarios, such as when consumers are faced with numerous product reviews or have little time constraints, they may decide to continue their search based on information they have acquired thus far (sequential search). Third, the assumption that consumers have perfect knowledge about how similar their preferences are to their social contacts can be relaxed. If consumers seek information from nascent relationships about an unfamiliar product category, they will learn about both similarity of preferences and their own product evaluation. Finally, an interesting extension would be to assess how consumers strategically categorize people into two or more groups (e.g., classifying contacts in Google Circles) and its impact on social learning.

## References

Ando, T. 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models". *Biometrika.* **94**(2) 443-458.

Camacho, N., B. Donkers, S. Stremersch. 2011. Predictably non-bayesian: quantifying sailence effects in physician learning about drug quality. *Marketing Science*, **30**(2) 305-320.

Carley, K. 1991. A theory of group stability. *American Sociological Review*. **56** 331-54.

Cartwright, D., F. Harary. 1956. Structural balance: A generalization of Heider's theory. *Psychological Review*. **63** 277-293.

Ching, A., T. Erdem, M. Keane. 2012. Learning models: An assessment of progress. Working Paper, University of Toronto.

De los Santos, B., A. Hortaçsu, M. Wildenbeest. 2012. Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review*. **102**(6) 2955-2980.

Erdem, T. 1998. An empirical analysis of umbrella branding. Journal of Marketing Research. **35**(3) 339-351.

Erdem, T., M. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*. **15**(1) 1-20.

Erdem, T. M. Keane, S. Öncü, J. Strebel. 2005. Learning about computers: An analysis of information search and technology choice. *Quantitative Marketing and Economics*. **3**(3) 207-246.

Godes, D., D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, P. Verlegh. 2005. The firm's management of social interactions. *Marketing Letters*. Volume **16**(3) 415-428

Hauser, J., G. Urban, B.Weinberg. 1993. How Consumers allocate their time when searching for information. *Journal of Marketing Research*. **30**(4), 452-466.

Heider, F. 1946. Attitudes and cognitive organization. *The Journal of Psychology.* **21**(1) 107–112.

Hartmann, W., P. Manchanda, H. Nair, M. Bothner, P. Dodds, D. Godes, K. Hosangar, C. Tucker. 2008. Modeling social interactions: Identification, empirical methods and policy implication. *Marketing Letters*. Volume **19**(3) 287-304.

Goethals, G., J. Darley. 1977. Social comparison theory: An attributional approach. *Social comparison processes: Theoretical and empirical perspectives*. Washington, DC: Hemisphere. 259-278.

Goethals, G., E. Nelson. 1973. Similarity in influence process: The belief-value distinction. *Journal of Personality and Social Psychology*. **25**(1) 117-122.

Hedesström, T., H. Svedsäter, and T. Gärling. 2006. Covariation neglect among novice investors. *Journal of Experimental Psychology*. **12**(3) 155–165.

Iyengar, R., K. Jedidi. 2012. A conjoint model for quantity discounts**.** *Marketing Science*. **31**(2) 334-350.

Levine, J., L. Resnick. 1993. Social foundation of cognition. *Annual Review of Psychology*. **44** 585-612.

Markowitz, H. Portfolio selection. *The Journal of Finance*. 7(1) 77-91.

McFadden, D. 1978. Modelling the choice of residential location. *Spatial Interaction Theory and Planning Models*. Amsterdam, North-Holland.

McPherson, M., L. Smith-Lovin, J. Cook. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*. **27** 415-444.

Mehta N., S. Rajiv, K. Srinivasan. 2004. The role of forgetting in memory-based choice decisions. *Quantitative Marketing and Economics*. **2**(2) 107-140.

Meyers-Levy, J., A. Tybout. 1989. Schema congruity as a basis for product evaluation. *Journal of Consumer Research*. **16**(1) 39-54.

Montoya, R., O. Netzer, K. Jedidi. 2010. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science*. **29**(5) 909–924.

Morgan, P., R. Manning. 1985. Optimal Search. *Econometrica*. **53**(4) 923 - 944.

Narayan, V., V. Rao, C. Saunders. 2011. How peer influence affects attribute preferences: A Bayesian updating mechanism. *Marketing Science*. **30**(2) 368-384.

Narayanan, S., P. Manchanda. 2009. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*. **28**(3) 424-441.

Neuberg, S., J. Newsom. 1993. Personal need for structure: Individual differences in the desire for simple structure. *Journal of Personality and Social Psychology*. **65**(1) 113-131.

Ratchford, B. 2010. Consumer Search and Pricing. *Handbook of Pricing Research in Marketing*, ed. Vithala R. Rao, 91–107. Northampton, MA: Edward Elgar Publishing.

Ratchford, B., M. Lee, D. Talukdar. 2003. The impact of the internet on information search for automobiles. *Journal of Marketing Research*. **40**(2) 193-209.

Roberts, J., G. Urban.1988. Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science*. **34**(2) 167-185.

Sinan, A., D. Walker. 2012. Identifying Influential and Susceptible Members of Social Networks. *Science.* **337** 337-341.

Spiegelhalter, D.. N. Best, B. Carlin, A. van der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B.* **64**(4) 583-639.

Stigler, G. 1961. The economics of information. *Journal of Political Economy.* **69**(3) 213-225

Suls, J., R. Martin, L. Wheele. 2000. Three kinds of opinion comparison: The triadic model. *Personality and Social Psychology Review*. **4**(3) 219-237.

Train, K., D. McFadden, M. Ben-Akiva. 1987. The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*. **18**(1) 109-123.

Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**(4157) 1124–1131.

Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*. **11** 3571–3594.

Weitzman, M. 1979. Optimal search for the best alternative. *Econometrica*. **47**(3) 641-654.

Winkler, R. 1981. Combining probability distributions from dependent information sources. *Management Science*. 27(4) 479-488.

Yaniv, I., S. Choshen-Hillelb, M. Milyavsky. 2011. Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes*. **115**(1) 111–120.

Zhao, Y., S. Yang, V. Narayan , Y. Zhao. 2012. Modeling consumer learning from online product reviews. *Marketing Science*. Forthcoming.

Zhao, Y., Y. Zhao, K. Helsen. 2011. Consumer learning in a turbulent Market environment: Modeling consumer choice dynamics after a product-harm crisis," *Journal of Marketing Research*. **48**(2) 255-267.

## Appendix 1. Regularity Conditions on Similarity Structure

We have two regularity conditions on the covariance matrix of triadic similarity – positive definiteness condition and inference condition.

*Positive-Definiteness Condition*. The covariance matrix for the triadic similarity structure should be a proper covariance matrix. Given the expression of similarity in Equation 5, we can write the positive-definiteness condition in the following way. (For notation simplicity, we omit subscripts for respondent and product)

$$1-a^2-b^2-c^2+2abc>0\,. \tag{A1}$$

*Inference Condition.* Intuitively, the information from more relevant sources should reduce the uncertainty more. As an extreme example, the information from those who have exactly same preference will completely resolve any uncertainty. The same is the case for information from those who have exactly the opposite preference. However, some similarity structures violate this intuition of learning process and uncertainty actually increases by acquiring information from more relevant sources. Suppose a similarity structure is represented by Equation 5 in the main text, and the respondent is informed of the values of $R^A$ and $R^B$. Then, the conditional distribution for his own evaluation, $R^E$, given $R^A$ and $R^B$ is:

$$f^{R^E}\left(R^E \mid R^A, R^B\right)=N\left(R^0+\frac{a-bc}{1-c^2}\left(R^A-R^0\right)+\frac{b-ac}{1-c^2}\left(R^B-R^0\right),\frac{\tau^2\left(1-a^2-b^2-c^2+2abc\right)}{1-c^2}\right). \tag{A2}$$

A similarity structure violates our intuition when the conditional variance, which denotes the uncertainty after having complete knowledge of $R^A$ and $R^B$, increases as relevance ($|a|$ or $|b|$) increases. As a result of comparative statics, we obtain the following condition for the conditional variance to not increase with the relevance of sources:

$$|a| \; > \; |bc| \text{ and } |b| \; > \; |ac|. \tag{A3}$$

We exclude similarity structures where the statistical axiom does not fit our intuition.

**Appendix 2. Informational Benefit under SMBL**

Posterior variance (*PV*) denotes the uncertainty that a consumer has about a product after search. (For notation simplicity, we omit the subscripts for respondent and product). That is, the posterior variance is closely associated with informational benefit of signals - the smaller is the posterior, the higher is the informational benefit. We can understand how the informational benefit from a signal is influenced by similarity structure under SMBL by examining the posterior variance. For ease of exposition, we can rewrite *PV* as a function of relevance (|*a*| and |*b*|) and redundancy (|*c*|) in the following way.

$$PV = \frac{\tau^2\left(\tau^2 n^A\left(1-|a|^2\right)+\sigma^{A^2}\right)\left(\tau^2 n^B\left(1-|b|^2\right)+\sigma^{B^2}\right)+\tau^6 n^A n^B\left(2D|abc|-|ab|^2-|c|^2\right)}{\left(\sigma^{A^2}+\tau^2 n^A\right)\left(\sigma^{B^2}+\tau^2 n^B\right)-|c|^2\tau^4 n^A n^B}, \tag{A4}$$

where *D* denotes an indicator of balance status which takes a value of 1 when *abc*>0 (balance), -1 when *abc*<0 (imbalance), and is 0 when *abc=0*.

*Sign of Relevance*. When the balance status (*D*) is held fixed, the sign of relevance per se does not affect the informational benefit. Suppose the relevance of a source (*a*) switches the sign, but balance status remains unchanged. We can think about a scenario where either *D*=0, or one of other similarity measures (*b* or *c*) also switches the sign when $D \neq 0$. Given that *a* enters Equation A4 only in absolute value, the sign of relevance per se will not affect informational benefit. The sign of relevance will change the informational benefit (*PV*) only through a change in the balance status of the system.

*Balance Status and Redundancy*. Given all the other values fixed, we can immediately see that PV is always greater when $D$=1 (balance) than when $D$=-1 (imbalance). That is, the informational benefit is always greater under imbalance than balance.

Next, the result of comparative statics shows that the effect of redundancy ($c$) on the posterior variance is contingent on the balance status ($D$).

$$\frac{\partial PV}{\partial |c|} = \frac{2D\tau^6 n^A n^B \left(|a|\sigma^{B^2} + \left(|a| - |bc|\right)\tau^2 n^A n^B\right)\left(|b|\sigma^{A^2} + \left(|b| - |ac|\right)\tau^2 n^A n^B\right)}{\left(\left(\sigma^{A^2} + \tau^2 n^A\right)\left(\sigma^{B^2} + \tau^2 n^B\right) - |c|^2 \tau^4 n^A n^B\right)^2}, \tag{A5}$$

Given the second regularity condition (A3), the sign of the first derivative is determined by balance status; it is positive when $D$=1, and negative when $D$=-1. In other words, the increase in redundancy (|$c$|) decreases (increases) the informational benefit under balanced (imbalanced) system.

## Appendix 3. Robustness Checks

*Subjective Average of Aggregate Evaluations.* We cannot jointly identify subjective average and subjective variance of aggregate evaluations (see model identification in Appendix 4), so we should fix either one to be the same as the manipulated value. In Model 8 (which we proposed as our main model), we fixed the subjective average to be same as the manipulated average, and estimated a parameter for the subjective variance $\left(\theta_{i0}^{pri}\right)$. To validate our specification, we fit an alternative specification where we fixed subjective variance to be the same as the manipulated variance $\left(\tau_{ij} = \tau_j^M\right)$, but allowed the subjective average to be different from what we manipulated $\left(R_{ij}^0 = \theta_{i0}^{pri} R_j^{0M}\right)$. All other components of this model are the same as Model 8. We found that this model was worse than our Model 8 specification on the hold-out sample fit (VLL=-1548.42). Also, there was no significant evidence that people perceive the average to be different from what we manipulated. In other words, subjectivity is associated more with a perception of variance than the average.

*Subjective Variance of Aggregate Evaluations.* In Section 5.2, we conceptually justified our specification where the subjective variance of aggregate evaluations does not depend on the structural aspects of preference similarity. In this section, we examine an alternative specification of our main model - subjective variation of aggregate evaluations is also specified as a function of the sign of relevance and balance status $\left(\tau_{ij} = \tau_j^M \times \exp\left(\theta_{i0}^{pri} + \theta_{i1}^{pri} neg_j + \theta_{i2}^{pri} imb_j\right)\right)$ where $neg_j$ denotes the number of negatively relevant sources. We found that hold-out sample fit was much worse than our main model specification (VLL=-1515.65). Neither of the two parameters, $\left(\theta_{i1}^{pri}, \theta_{i2}^{pri}\right)$, was significant and all the substantial findings were robust. In other words, subjective variance of aggregate ratings was not affected by similarity structure.

**Appendix 4. Identification of Population Parameters**

In this section, we outline the identification of population parameters. Given the identification of population parameters $(\Gamma, \Delta, \Theta, \lambda)$, one can easily identify individual-level parameters $(\Gamma_i, \Delta_i, \Theta_i, \lambda_i)$ with distributional assumptions on individual parameters (i.e., normally distributed around population parameters).

First, we can identify $\Gamma$ and $\Theta$ from observed purchase decisions given the search decisions. Observed purchase decisions with $n_{ij}^* = (0,0)$ can identify utility parameters $\left(\Gamma = [\gamma_0, \beta]\right)$ and perceived prior variance $\left(\theta_0^{pri}\right)$.[1] Given that there was no search at all, the general tendency of purchase is captured by $\gamma_0$, the effect of manipulated prior mean $\left(R_j^{0M}\right)$ on purchase is captured

[1] In our data, we have 147 observations with no search at all, and 13 of them converged to purchase.

by $\beta$, and the effect of manipulated prior standard deviation $\left(\tau_j^M\right)$ on purchase is captured by $\beta$ and $\theta_0^{pri}$.

The parameters related to perceived similarity with each sources $\left(\theta_0^{rel},\theta_1^{rel}\right)$ are identified from observed purchase decisions given (1) no search from one source and (2) a sufficiently large amount of search from the other source (i.e., steady state where an additional signal hardly increase the search utility).[2] For these observations, purchase utility depends only on parameters identified above (i.e., utility parameters and perceived prior variance) and perceived similarity with the source where the steady state is achieved ($a_j$ or $b_j$). Therefore, general tendency of purchase among these observations identifies parameter $\theta_0^{rel}$, and the difference in purchase driven by the sign of relevance identifies $\theta_1^{rel}$.

The parameters related to perceived similarity between the two sources $\left(\theta_0^{red},\theta_1^{red}\right)$ are identified from the observed purchase decisions given a sufficiently large amount of search from both sources.[3] For these observations, purchase utility depends only on parameters identified so far (i.e., utility parameters, perceived prior variance, and perceived similarity with each source) and perceived similarity between the sources ($c_j$). Therefore, general tendency of purchase among these observations identifies parameter $\theta_0^{red}$, and the difference in purchase driven by balance status identifies the parameter $\theta_1^{red}$.

[2] Our model estimates suggest that average perceived signal variance is around 0.12 and average perceived prior variance is around 0.15. Given this information, the steady state is roughly achieved when the search amount is greater than 10. In our data, 45 observations reached steady state for one source, but did no search from the other source. Among those observations, 30 observations collected signals from positively relevant source only (15/30 converged to purchase), and 15 observations collected signals from negatively relevant source only (1/15 converged to purchase). Therefore, we have sufficient information to identify ($\theta_0^{rel}$, $\theta_1^{rel}$).

[3] In our data, around 288 observations reached steady state for both sources. Among those observations, 142 observations were under balance (50/142 converged to purchase), and 146 were under imbalance (46/142 converged to purchase). Therefore, we have sufficient information to identify ($\theta_0^{red}$, $\theta_1^{red}$).

The increase in purchase likelihood with respect to the amount of observed search ( $n_{ij}^{*}$ ) will identify the parameters of the perceived signal variance $\left(\theta_0^{sig}, \theta_1^{sig}, \theta_2^{sig}\right)$. Specifically, a general pattern of increase in purchase likelihood identifies $\theta_0^{sig}$. The parameter $\theta_1^{sig}$ is identified from any systematic differences in the purchase pattern when signals are collected from a source with a negative relevance as opposed to a source with a positive relevance. Similarly, the parameter $\theta_2^{sig}$ is identified from any systematic differences in the purchase pattern under an imbalanced system as opposed to a balanced system.

Finally, we can identify cost-related parameters $(\Delta)$ and scale parameter $(\lambda)$ with observed search decisions. Given the identification of $\Gamma$ and $\Theta$, the expected informational benefit, $v_j^b(\cdot)$ in Equation 19, is identified. The effect of the expected informational benefit on search decision will identify $\lambda$. The effect of $\left(n_{ij}^A, n_{ij}^B\right)$ on the search decision captured through $v_j^c(\cdot)$ will identify $\delta_0$. The parameter $\delta_1$ is identified from a systematic difference in $v_j^c(\cdot)$ when signals are collected from a source with a negative relevance as opposed to a source with a positive relevance. Similarly, the parameter $\delta_2$ is identified from any systematic difference in $v_j^c(\cdot)$ when signals are collected under an imbalanced system as opposed to a balanced system.