

NET Institute\*

[www.NETinst.org](http://www.NETinst.org)

Working Paper #07-10

Revised April 2008

**Blockbuster Culture's Next Rise or Fall:  
The Impact of Recommender Systems on Sales Diversity**

Daniel Fleder and Kartik Hosanagar  
The Wharton School, University of Pennsylvania

\* The Networks, Electronic Commerce, and Telecommunications (“NET”) Institute, <http://www.NETinst.org>, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, “virtual networks” comprised of computers that share the same technical standard or operating system, and on network issues in general.

Blockbuster Culture's Next Rise or Fall:  
The Impact of Recommender Systems on Sales Diversity

Daniel Fleder and Kartik Hosanagar<sup>†</sup>  
The Wharton School, University of Pennsylvania  
{dfleder,kartikh}@wharton.upenn.edu

Draft: April, 2008

This paper examines the effect of recommender systems on the diversity of sales. Two anecdotal views exist about such effects. Some believe recommenders help consumers discover new products and thus increase sales diversity. Others believe recommenders only reinforce the popularity of already popular products. This paper seeks to reconcile these seemingly incompatible views. We explore the question in two ways. First, modeling recommender systems analytically allows us to explore their path dependent effects. Second, turning to simulation, we increase the realism of our results by combining choice models with actual implementations of recommender systems. Our main result is that some well known recommenders can lead to a reduction in sales diversity. Because common recommenders (e.g., collaborative filters) recommend products based on sales and ratings, they cannot recommend products with limited historical data, even if they would be rated favorably. In turn, these recommenders can create a rich-get-richer effect for popular products and vice-versa for unpopular ones. This bias toward popularity can prevent what may otherwise be better consumer-product matches. That diversity can decrease is surprising to consumers who express that recommendations have helped them discover new products. In line with this, we show it is possible for individual-level diversity to increase but aggregate diversity to decrease. Recommenders can push each person to new products, but they often push similar users toward the same products. We show how basic design choices affect the outcome, and thus managers can choose recommender designs that are more consistent with their sales goals and consumers' preferences.

---

<sup>†</sup> The authors thank Yannis Bakos, Eric Bradlow, Terry Elrod, Pete Fader, Robin Pemantle, David Schweidel, Michael Steele, Christophe Van den Bulte, and seminar participants at Carnegie Mellon University, Stanford University, University of Connecticut, University of Pennsylvania and WISE for their comments. This work is a much extended version of a conference paper by the same authors (2007) from the ACM Conference on Electronic Commerce, and we thank their three anonymous reviewers. The authors also thank Barrie Nault, an associate editor, and two anonymous reviewers for their valuable suggestions. The paper benefited thoroughly from their comments, and any remaining errors are the authors'. Finally, we thank we thank the NET Institute ([www.NETinst.org](http://www.NETinst.org)) for financial support.

## 1. INTRODUCTION

Media has historically been a “blockbuster” industry. Of the many products available, sales concentrate among a small number of hits. In recent years, such concentration has begun to decrease. The last ten years have seen an extraordinary increase in the number of products available (Brynjolfsson et al. 2006; Clemons et al. 2006), and consumers have taken to these expanded offerings. It is believed this increased variety will allow consumers to obtain the ideal products for them, and if it continues could amount to a cultural shift from hit to niche products. One difficulty that arises, however, is how consumers will find such niche products among seemingly endless alternatives.

Recommender systems are considered one solution to this problem. These systems use data on purchases, product ratings, and user profiles to predict which products are best suited to a particular customer. These systems are commonplace at major online firms such as Amazon, Netflix, and Apple’s iTunes Store. In author Chris Anderson’s view, “The main effect of filters, [which include online recommender systems], is to help people move from the world they know (‘hits’) to the world they don’t (‘niches’)” (2006, p. 109).

Will recommenders make us all viewers of niche, independent movies and music? Or, might they actually reinforce the blockbuster nature of media? While recommenders have been assumed to push consumers toward the niches, we present an argument why some popular systems might do the opposite.<sup>2</sup> Anecdotes from users and researchers suggest recommenders help consumers discover new products, and thus increase sales diversity (Anderson 2006). A small number of others believe several recommender designs have the potential to lower diversity by reinforcing the position of already popular products (Mooney & Roy 2000; Fleder & Hosanagar 2007). This paper attempts to reconcile these seemingly incompatible views. Holding supply-side offerings fixed, we ask whether recommenders make media consumption more diverse or more concentrated.

---

<sup>2</sup> With so many different recommenders employed by firms, one cannot state a universal result for all. Instead this paper picks several recommenders we believe are commonly used in industry and focuses on them.

We explore this question in two ways. First, modeling recommender systems analytically allows us to explore their path dependent effects. Second, using simulation, we increase the realism of our results by combining choice models with actual implementations of recommender systems. Our main result is that some popular recommenders can lead to a reduction in diversity. Because common recommenders (e.g., collaborative filters) recommend products based on sales or ratings, they cannot recommend products with limited historical data, even if they would be rated favorably. These recommenders create a rich-get-richer effect for popular products and vice-versa for unpopular ones. Several popular recommenders explicitly discount popular items, in an effort to promote exploration. Even so, we show this step may not be enough to increase diversity.

That diversity can decrease is surprising to consumers who express that recommendations have helped them discover new products. The model provides two insights here. First, we find it is possible for individual-level diversity to increase but aggregate diversity to decrease. Recommenders can push each person to new products, but they often push similar users toward the same products. Second, if recommenders are replacing best-seller lists, diversity can increase by cutting out an even more popularity-biased tool; whereas relative to a world with neither, recommenders may lower diversity.

The results have implications for firms and consumers. For retailers, we show how design choices affect sales and diversity. For consumers and niche content producers, we show how a recommender's bias toward popular items can prevent what would otherwise be better consumer-product matches. We find that recommender designs that explicitly promote diversity may be more desirable.

The rest of the paper is organized as follows. Section 2 reviews prior work. Section 3 gives a formal problem statement. Section 4 presents the analytic model, which is stylized but still able to show how sales information can bias recommenders. To increase the realism of our setting, and in particular incorporate actual recommender designs, a complementary simulation is developed in Sections 5-7. The simulation combines consumer choice models with actual recommender algorithms. Section 8 discusses

the implications for producer and consumer welfare. Section 9 concludes, reviewing the findings and offering directions for future work.

## **2. PRIOR WORK**

Recommender systems help consumers learn of new products and select desirable products among myriad choices (Pham & Healey 2005). These systems can be seen as helping to automate the word of mouth process (Shardanand & Maes 1995; Resnick & Varian 1997; Dellarocas 2003). A simplified taxonomy divides recommenders into content-based versus collaborative filter-based systems. Content-based systems use product information (e.g., author, genre, mood) to recommend items similar to those a user rated highly. Collaborative filters, in contrast, recommend what similar customers bought or liked. Perhaps the best-known collaborative filter is that seen on Amazon.com, with its familiar tagline, “Customers who bought this item also bought... .” The current work focuses on collaborative filters, which appear to be more common in industry.

The design of these systems is an active research area. Reviews are provided in Breese et al. (1998) and Adomavicius and Tuzhilin (2005). For business contexts, Ansari et al. (2000) describes how firms can integrate other data sources (e.g., expert opinions) into recommendations. Work by Bodapati (2008) places recommender systems into a profit-maximizing framework. For industry applications, implementations at firms such as Amazon.com and CDNOW are described by Schafer et al. (1999), Sarwar et al. (2001), and Linden et al. (2003). While there is a large body of work on building these systems, we know much less about how they affect consumer choice and behavior.

Studies have recently begun to examine individual-level, behavioral effects. In marketing, Senecal and Nantel (2004) show experimentally that recommendations do influence choice. They find that online recommendations can be more influential than human ones. Cooke et al. (2002) examine how purchase decisions under recommendations depend on the information provided, context, and familiarity.

While the above studies ask how recommenders affect individuals, our interest is the aggregate effect they have on markets and society. In particular, we are interested in their effect on sales diversity. To the

best of our knowledge, there have not been formal studies isolating such effects, although the topic has received mention from several researchers. Brynjolfsson et al. (2007) find that a firm's online sales channel has slightly higher diversity than its offline channel. They suggest demand-side causes, such as active tools (search engines) and passive tools (recommender systems), but do not isolate the specific effect of recommenders. In contrast, Mooney and Roy (2000) suggest collaborative filters may perpetuate homogeneity in choice, but it is an in-passing comment without formal study.

Given our focus on aggregate effects, we would also like to highlight two related streams of work, namely those on information cascades and Internet Balkanization. The information cascades literature has looked at aggregate effects of observational learning and resulting convergence in behavior, or "herding" (Bikhchandani et al. 1998). The Internet balkanization literature has studied whether the Internet helps create a global community freed of geographic constraints. Van Alstyne and Brynjolfsson (2005) find that while increased integration or diversity can result, the Internet can also lead to greater Balkanization wherein groups with similar interests find each other and become more homogeneous. While our problem is different, we see these papers as complementary in highlighting the social implications of technologies that share information among users.

This prior work reveals four themes. One, recommender systems research in the data mining literature has focused more on system design than understanding behavioral effects. Two, the marketing literature is just beginning to examine such behavioral effects. Three, of the existing behavioral work, the focus has been more on individual outcomes than aggregate effects. Four, regarding aggregate effects, there are opposing conjectures as to the effect of recommenders on sales diversity.

### **3. PROBLEM DEFINITION**

This section sets the problem context, defines a measure of sales diversity, and formulates the question to be investigated. Our context is a market with a single firm selling one class of good (e.g., music versus movies). Within this one class, the firm can offer many items (e.g., CDs by thousands of artists).

### 3.1 Measure of Sales Diversity

Before examining recommender systems' effects, it is necessary to distinguish between sales and product diversity. *Product diversity*, or product variety, typically measures how many different products a firm offers. It is a supply-side measure of breadth. In contrast, we use *sales diversity* to describe the concentration of market shares conditional on firms' assortment decisions. To measure sales diversity, we adopt the Gini coefficient. The Gini coefficient is a common measure of distributional inequality and has been applied to many problems, the most common being perhaps wealth inequality. For additional discussion of this measure, see Sen (1976); for usage related to recommendations, see Oestreicher-Singer & Sundararajan (2006).

Let  $L(u)$  be the Lorenz curve denoting the percentage of the firm's sales generated by the lowest  $100u\%$  of goods sold during a fixed time period. The Gini coefficient is defined  $G := 1 - 2 \int_0^1 L(u) du$ .

Graphically this corresponds to  $G = A/(A+B)$  in Figure 1. Thus  $G \in [0,1]$ , and it measures how much the Lorenz curve deviates from the  $45^\circ$  line. A value  $G = 0$  reflects diversity (all products have equal sales), while values near 1 represent concentration (a small number of products account for most of the sales).

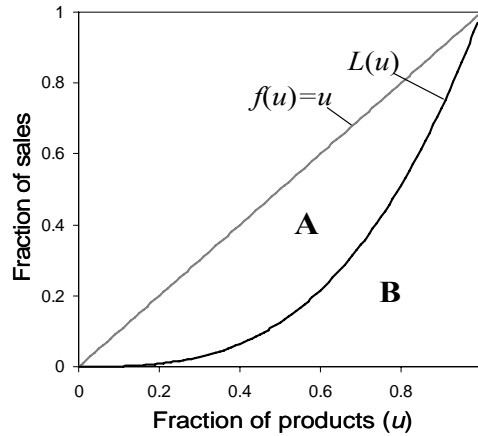


Figure 1. Lorenz curve

### 3.2 Problem Statement

Consider a firm with  $I$  customers  $c_1, \dots, c_I$  and  $J$  products  $p_1, \dots, p_J$ . Define a recommender system as a function  $r$  that maps a customer  $c_i$  and database onto a recommended product  $p_j$ . Typically the database

records consumer purchases and/or ratings data. Consider next a set of different recommender systems  $r_1, \dots, r_k$ . Each  $r_i$  reflects certain design choices. For example,  $r_i$  might be a user-to-user collaborative filter, while  $r_j$  might be Amazon.com's item-to-item collaborative filter. Denote by  $G_0$  the Gini coefficient of the firm's sales during a fixed time period in which a recommender system was *not* used. In contrast, let  $G_i$  be the Gini coefficient of the firm's sales during a period in which recommender system  $r_i$  was employed with all else equal.

**Definition.** *Recommender bias.* Recommender  $r_i$  is said to have a concentration bias, diversity bias, or no bias depending on the following conditions:

$$\left\{ \begin{array}{ll} \text{Concentration bias} & G_i > G_0 \\ \text{Diversity bias} & G_i < G_0 \\ \text{No bias} & G_i = G_0 \end{array} \right.$$

For various recommenders, we examine whether a bias exists and its direction.

#### 4. ANALYTICAL MODEL

This section presents an analytical model to explore a recommender system's effects. We present a stylized model of a collaborative filter. The simple model illustrates how the use of sales information by recommender systems affects diversity. In Section 5, we consider more realistic specifications for the recommender and the consumer's decision process. Throughout the paper, recommender system is synonymous with collaborative filter. Collaborative filters can operate on purchase or ratings data. To fix a context, our model considers purchases.

##### 4.1 Assumptions and Model

We consider a set of customers making purchases sequentially. This section enumerates the assumptions defining the analytic model.

**Assumption 1.** *Each consumer buys one product per time step.* The customer's decision is which product to buy and not whether to buy. For example, at a streaming media service, this could reflect customers who have decided to listen to a personalized radio station for an hour and whose playlists are determined by the recommender. This assumption helps isolate choice from purchase incidence.



**Assumption 2.** We assume there are only two products,  $w$  and  $b$  (white and black). This assumption is for tractability.

**Assumption 3.** Consumers have purchase probabilities  $(p, 1-p)$  for  $(w, b)$  in the absence of recommendations. We do not model the decision process that generates these purchase probabilities.

**Assumption 4.** At each occasion, the firm recommends a product, which is accepted with probability  $r$ . Thus,  $r$  is the strength of the recommender.

**Assumption 5.** The firm's recommendation is generated using a function  $g(X_t) \in \{w, b\}$ , where  $X_t$  is the segment share of  $w$  just before purchase  $t$ . The recommender's inputs are segment shares, market shares within a particular segment of similar users, and its output is a product recommendation. The system modeled recommends the product with higher segment share. This choice of  $g$  has a parallel with collaborative filters. Many collaborative filters find similar customer segments and recommend the most popular item within them (e.g., "people who bought  $X$  also bought  $Y$ "). This recommender can be represented by the step function

$$g(X_t) := P(w \text{ recommended} | X_t) = \begin{cases} 0 & , X_t < \frac{1}{2} \\ \frac{1}{2} & , X_t = \frac{1}{2} \\ 1 & , X_t > \frac{1}{2} \end{cases} \quad (1)$$

where  $X_t \in [0, 1]$ . Figure 3 plots this. When  $X_t = \frac{1}{2}$  and the products have equal shares, the recommendation is determined by a Bernoulli( $\frac{1}{2}$ ) trial. To start, the recommender does not favor either product, and we assume  $X_1 = \frac{1}{2}$ .

**Assumption 6.** The segment of consumers constituting  $X_t$  is pre-selected and does not change over time. This segment of similar consumers is identified based on past behavior, possibly from purchases of products in other categories. The assumption that the group does not evolve is for tractability, since such sequential user similarity (nearest-neighbor) calculations are difficult to model analytically. This assumption is for simplicity, but it does have a parallel with business practice. In industry, real-time updating of segments is often computationally prohibitive, and so many firms update segments periodically. Section 5 presents an alternate approach where we let the segments evolve over time.

The process defined by these assumptions can be illustrated by an urn model. Urn models are appealing for stochastic processes, and Johnson and Kotz (1977) show how many significant results from probability theory can be derived from such settings. Consider the two urn system of Figure 2. Urn 1 contains balls representing products  $w$  and  $b$ . A fraction  $p$  of the balls in urn 1 are white; this fraction is the consumer's purchase probability for  $w$  in the absence of recommendations. Urn 2 is the recommender: its contents reflect the sales history within the segment, and it produces recommendations according to  $g(X_t)$ , where  $X_t$  is the fraction of  $w$  in urn 2 just before  $t$ . To start, urn 2 contains one  $w$  and one  $b$ . At time  $t=1$ , a ball is drawn *with* replacement from urn 1 representing the consumer's choice before seeing the recommendation. Next, a ball is drawn *with* replacement from urn 2 according to  $g(X_t)$ , representing the recommended product. With probability  $r$ , the consumer accepts the recommendation, and with probability  $(1-r)$  the consumer retains the original choice. The ball chosen represents the actual product purchased; afterward a copy of it is added to urn 2, which is equivalent to updating the recommender's sales history (e.g., the firm's database). Consumer 2 then arrives, and the process repeats ( $p$  and  $r$  are the same, but  $X_2$  is used instead of  $X_1$ ), and so on for other customers.

From these assumptions, the probability that  $w$  is purchased at time  $t$  is

$$\begin{aligned}
 f(X_t) &:= P(w \text{ chosen on occasion } t \mid X_t) \\
 &= p(1-r) + g(X_t)r \\
 &= \begin{cases} p(1-r) & , X_t < \frac{1}{2} & \text{"l"} \\ \frac{[p(1-r)] + [p(1-r) + r]}{2} & , X_t = \frac{1}{2} & \text{"m"} \\ p(1-r) + r & , X_t > \frac{1}{2} & \text{"h"} \end{cases} \quad (2)
 \end{aligned}$$

Figure 4 plots an example of  $f$ . The labels in (2) "l", "m", "h" are short-hand; they visually refer to the low ( $l$ ), middle ( $m$ ), and high ( $h$ ) portion of  $f$ 's shape in Figure 4. The geometry of this figure helps illustrate the results derived next.

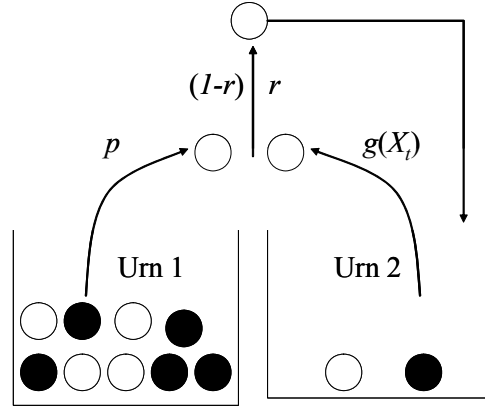


Figure 2. A two-urn model for recommender systems

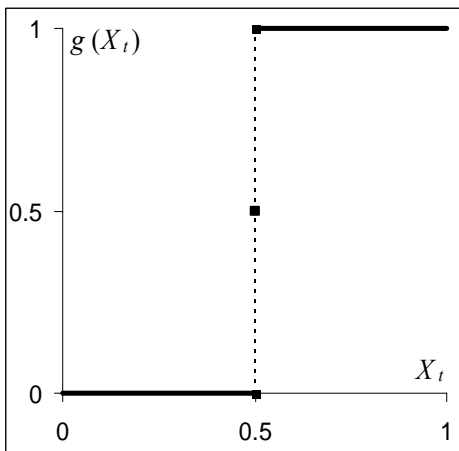


Figure 3. Recommender  $g(X_t)$

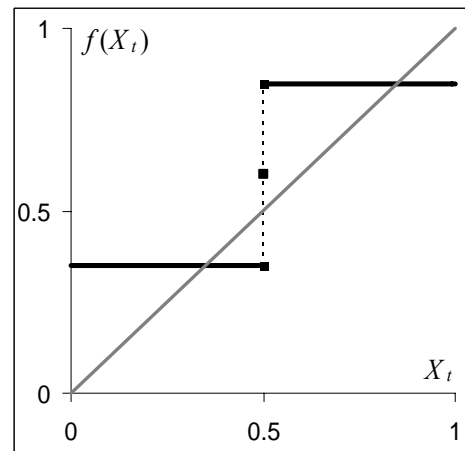


Figure 4. Urn function  $f(X_t)$  and 45° line ( $p=0.7, r=0.5$  example)

## 4.2 Model Results

Results are given in three parts. First we derive properties of market shares under recommendations. Second we show a graphical example. Third, we relate these market shares to the question of diversity.

### 4.2.1 Theoretical results

The following results are derived in a random walks framework by examining the difference  $w - b$  over time. For clarity, all proofs are in the appendix.

Without recommendations, shares converge to  $(p, 1-p)$ . The first proposition asks how this is affected by the presence of a recommender. As  $t \rightarrow \infty$ ,  $\{X_t\}$  will converge to one of two values. These limiting values depend on the consumer's initial  $p$  and recommender's influence  $r$  and are given by

**Proposition 1.** *Support points. As  $t \rightarrow \infty$ ,  $X_t$  converges to*

Case	Support point 1	Support point 2
1. $p \leq (\frac{1}{2} - r)/(1 - r)$	$l$	n/a
2. $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r)$	$l$	$h$
3. $p \geq \frac{1}{2}/(1 - r)$	$h$	n/a

where the shorthand  $l$  and  $h$  are from equation (4),  $p \in [0,1]$ , and  $r \in (0,1)$  ( $r = 0$  or  $1$  is trivial).

The cases in Proposition 1 have an attractive geometric interpretation: The support points are simply the intersections of  $f(X_t)$  with the 45° line in Figure 4. That is, the support points are  $\{x : f(x) = x\}$ .<sup>3</sup> Visually,  $p$  and  $r$  shift and stretch the step function; as a result, it has either one intersection occurring below  $f(X_t) = 0.5$  (Case 1), one intersection occurring above  $f(X_t) = 0.5$  (Case 3), or both (Case 2).

**Corollary 1.** *Chance and winning the market. In Case 2,  $P(\lim_{t \rightarrow \infty} X_t < 1/2) > 0$  and  $P(\lim_{t \rightarrow \infty} X_t > 1/2) > 0$ .*

This is evident because  $l < 0.5$  and  $h > 0.5$  are both support points. This shows an interesting aspect of Case 2: regardless of the initial  $p$ , either product can obtain and maintain the majority share.

With the limiting value(s) of  $\{X_t\}$  known, we ask whether they reflect higher or lower concentration. Let the term “increased concentration” define shares that are less equal than they would be without recommendations. Increased concentration means  $\lim_{t \rightarrow \infty} X_t > p$  when  $p > 1/2$  and  $\lim_{t \rightarrow \infty} X_t < p$  when  $p < 1/2$ . The effect on concentration is given by the following proposition.

**Proposition 2.** *Relation of limit points to concentration. For any  $(p,r)$ , the effect on concentration is*

Case	Support points	Effect on concentration relative to $p$
1	1	Increased concentration
2	2	Case 2A: $p \in (\frac{1-r}{2-r}, \frac{1}{2-r})$ . Increased concentration for both support points Case 2B: $p \notin (\frac{1-r}{2-r}, \frac{1}{2-r})$ . Increased concentration for one support point; decreased for the other
3	1	Increased concentration

These cases are shown visually in Figure 5. For Cases 1 and 3, there is a single outcome and that outcome always has increased concentration. These are areas of the  $p \times r$  space where consumers have

<sup>3</sup> The visual interpretation applies only to where  $f$ 's line segments intersect the 45° line (not the single point at  $X_t = 0.5$ ).

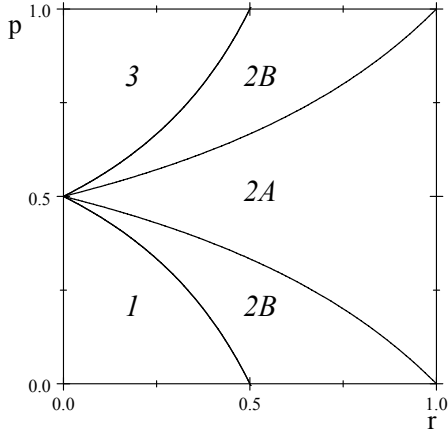
strong initial probability ( $p$ ) relative to the recommender's strength ( $r$ ); as a result, the recommender's effect is to reinforce this tendency even more. For example, if consumers have a fairly strong tendency to buy  $w$  with  $p = .90$  and the recommender is fairly influential with  $r = .25$ , the recommender creates a positive feedback loop, reinforcing the popularity of  $w$  and giving it a limit share of  $0.93 > 0.90$ . Product  $w$  was initially bought more, which made it recommended more, which made it bought more, and so on.

Case 2A occurs where the recommender's influence ( $r$ ) is high relative to the initial probability ( $p$ ). This has two implications, one at the sample-path and one at the aggregate level. At the sample path level, *either* product can win the market, regardless of  $p$ . For example,  $p = 0.55$  and  $r = 0.75$  imply limiting market shares of  $(w,b) \in \{(0.89,0.11), (0.14,0.86)\}$ . In the first outcome,  $w$  wins the market. In the second,  $b$  wins, even though  $p = 0.55$  initially favored  $w$  (c.f. Corollary 1). This occurs because  $r$  is large relative to  $p$ , and the recommender reinforces whichever product does well early on without too much resistance from  $p$ . This leads to the finding that recommenders can *create hits*. Some product will become a winner with a permanent, majority share, but we cannot say which beforehand. At the aggregate level, concentration always increases. We do not know which of  $w$  or  $b$  will win, but we know that one will and whichever does will be an outcome with greater concentration.<sup>4</sup>

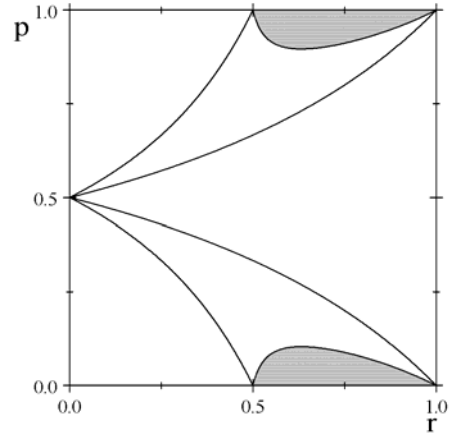
Last, in Case 2B, neither the initial probability ( $p$ ) nor the recommender's influence ( $r$ ) are strong relative to one another. As a result, two outcomes are possible. The tendency  $p$  can be reinforced by the recommender. This increases concentration. Or, the recommender can give whichever product was not favored a small majority. This decreases concentration. For example, if  $p = .60$ , which is a mild preference for  $w$ , and  $r = .25$ , which is low, the limit points are  $.70$  and  $.45$ . Often  $w$  has more early successes and the recommender reinforces this, leading to the  $.70$  outcome. In some cases, if  $b$  is chosen enough early on, the recommender reinforces  $b$  leading to the  $.45$  outcome, which entails less concentration than the initial share of  $.40$ .

---

<sup>4</sup> Although they start with different models, a similar phenomenon occurs in other contexts (e.g., studies of firm location). Arthur (1994) provides an overview of applications, while several underlying mathematical results are in Hill (1980)



**Figure 5. Relating the  $p \times r$  space to concentration effects (numbers refer to cases in Proposition 2).**



**Figure 6. Concentration increases in white areas and decreases in shaded ones.**

While both outcomes are possible in 2B, they are not equally likely. Next we determine the probability of arriving at each. This, in turn, allows us to calculate the expected effect on concentration.

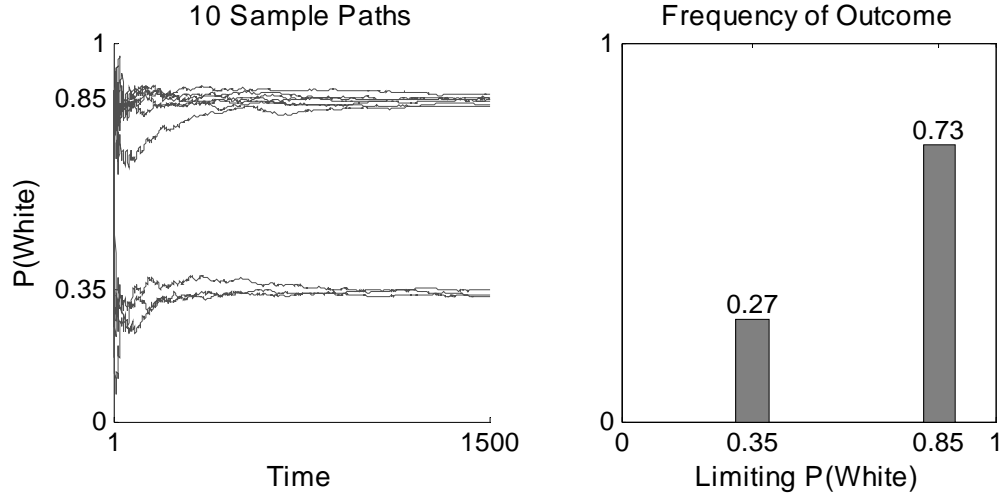
**Proposition 3.** *The distribution of  $\lim_{t \rightarrow \infty} X_t$  is*

Case	Support point 1	Support point 2	$P(\lim_{t \rightarrow \infty} X_t = \text{support point 1})$	$P(\lim_{t \rightarrow \infty} X_t = \text{support point 2})$
1. $p \leq (\frac{1}{2} - r)/(1 - r)$	$l$		1	0
2. $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r)$	$l$	$h$	$\gamma$	$1 - \gamma$
3. $p \geq \frac{1}{2}/(1 - r)$	$h$		1	0

where  $\gamma = \frac{(1-m) \cdot (1 - \frac{l}{1-l})}{m \cdot (1 - \frac{1-h}{h}) + (1-m) \cdot (1 - \frac{l}{1-l})} \in (0,1)$ . This proposition will be applied subsequently.

#### 4.2.2 Graphical example

A graphical example helps illustrate the results. For sake of illustration, take  $p = .70$  and  $r = .50$ . Figure 7 plots 10 realizations of this process over time. The left part of the figure shows these paths converging to two outcomes. Visually, one sees the limits are in accord with Proposition 1, which says the process converges to a random variable whose support is  $\{0.35, 0.85\}$ . At right, the figure shows that the frequencies of arriving at the lower versus upper outcome approach .27 and .73 respectively which is in accord with Proposition 3.



**Figure 7. The two limiting outcomes for our example  $f(x)$**

#### 4.2.3 Net effect on sales concentration

With the limiting distribution of  $\{X_t\}$  known, we complete the connection to sales concentration. For two products with shares  $p$  and  $1-p$ , the Gini coefficient is proportional to (Sen 1976)

$$G(p) = |p - \frac{1}{2}| \quad (3)$$

With recommendations, we define

$$\begin{aligned} G_{p,r} &= E[G(\lim_{t \rightarrow \infty} X_t) \mid p, r] \\ &= G(l)P(\lim_{t \rightarrow \infty} X_t = l) + G(h)P(\lim_{t \rightarrow \infty} X_t = h) \end{aligned} \quad (4)$$

The net effect on concentration is given by  $G_{p,r} - G_{p,0}$ , which is  $>0$  ( $<0$ ) when concentration increases (decreases). Substituting into (3) and (4) terms from the previous propositions gives

Case	$G_{p,r}$	$G_{p,0}$
1 $p \leq (\frac{1}{2} - r)/(1-r)$	$ l - \frac{1}{2} $	$ p - \frac{1}{2} $
2 $(\frac{1}{2} - r)/(1-r) < p < \frac{1}{2}/(1-r)$	$ l - \frac{1}{2} \gamma +  h - \frac{1}{2} (1-\gamma)$	$ p - \frac{1}{2} $
3 $p \geq \frac{1}{2}/(1-r)$	$ h - \frac{1}{2} $	$ p - \frac{1}{2} $

The above gives a closed-form expression for the change in Gini coefficient. Figure 6 shows this visually. For most of the  $p \times r$  square, concentration increases. This is true, of course, for areas under Case 1, 2A, and 3, where the only possibility was increased concentration. It is also true for most areas where both outcomes were possible (Case 2B). In extreme cases, it is possible for a net decrease to occur, as shown

by the shading. These areas are largely an artifact of the initial conditions assumed for urn 2, which place one  $w$  and one  $b$  in a high  $r$  recommender even when  $p \approx 0$  or  $\approx 1$ .<sup>5</sup>

We now summarize the findings of this section. Under recommendations, the shares converge to either one or two limiting outcomes depending on  $(p,r)$ . When there is one outcome, it always reflects increased concentration; the recommender reinforces the popularity of the initially preferred product. In the two outcome cases, either both outcomes have greater concentration or one has greater concentration and the other has less. For the latter, a net effect must be calculated. This typically has greater concentration, although for extreme  $(p,r)$ , as discussed, increased diversity may occur. Thus the recommender seems to increase concentration among a set of similar users.

The conclusions of this section are based on a stylized model of recommenders in a context with two products and a set of similar users that are aware of both products. To increase their realism, we now use simulation to combine multi-product choice models with actual recommender system implementations.

## 5. SIMULATIONS

### 5.1 Rationale for Simulation

Simulation offers three benefits for this problem. First, while actual recommender algorithms are difficult to represent analytically, they can be easily implemented in simulation. Second, a challenge in analytic models is heterogeneity. Heterogeneity can enter in two ways: users can have different preferences over products and the recommender can respond differently to each user. Both can be accommodated in simulation. Third, more complex choice processes can be represented than in the urn specification.

### 5.2 Choice Model and Simulation Design

We investigate the sales diversity question by using a simulation that combines a choice model with actual recommender systems. We assume the number of products supplied is fixed and that repeat

---

<sup>5</sup> An example illustrates how this is related to initial conditions. Suppose  $p = 0.99$ , and  $r = .99$ , which is in the shaded region. Since  $X_1 = .50$ ,  $P(b \text{ on first purchase}) \approx .50$ . If  $b$  is chosen, the recommender next suggests  $b$ ; since  $r = .99$  the next consumer is almost certain to pick  $b$  too, and so on for the remaining consumers even though  $p = 0.99$  favors  $w$ . We have conducted experiments where the initial conditions are determined by  $k$  Bernoulli( $p$ ) trials. In such cases, the shaded areas begin to fill in for even small  $k$ . (These additional experiments are available from the authors on request.)



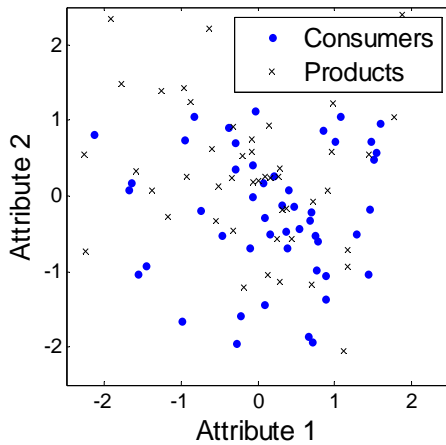
purchases are permitted. Examples of contexts with repeat purchases could include music and video streaming from a subscription service (e.g., Rhapsody).

An overview of the process is as follows. There are  $I$  consumers and  $J$  products positioned in an attribute space. Consumers are not aware of all products. Each consumer knows most of the central products and a small number of products in his own neighborhood. Every period, a consumer either purchases one of the products or makes no purchase at all. To model this choice, a multinomial logit is used for  $J$  products plus an outside good. Just before choosing a product, a recommendation is generated. The recommender has two effects. First, the consumer becomes aware of the recommended product if he was not already. This increase in awareness is permanent. Second, the salience of the recommended product is raised temporarily, increasing the chance that the recommended product is purchased. The next consumer makes a purchase in a similar manner, and the process repeats after all consumers have purchased. After a predetermined number of iterations, the Gini is computed. The Gini is then compared to a benchmark  $G_0$ , the Gini from an equivalent period in which recommendations were not offered.

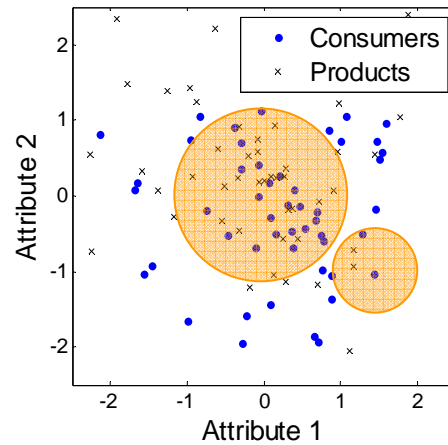
We now discuss each of the simulation components: (i) the map of product and consumer points, (ii) the recommender system  $r$ , (iii) the awareness distribution, (iv) the choice model, and (v) the salience factor  $\delta$ .

(i) *Map of product and consumer points.* The map of products and ideal points is the input for the choice model. Plotting consumer points and product locations goes back at least to Hotelling (1929) and is commonly used in marketing (e.g., Elrod & Keane 1995). Our consumers and products are points in a two-dimensional space. The use of two dimensions is for simplicity and visualization; for contexts with more than two attributes, the maps can be considered dimensionality-reduced versions, as is common in marketing research. We take both ideal points and products to be standard multivariate normal. The normality assumption for consumers is common in factor-analytic market maps (e.g., Elrod 1988; Elrod

& Keane 1995). Our base case uses 50 consumer and 50 product points, an example of which is shown visually in Figure 8.<sup>6</sup>



**Figure 8. Map of product and consumer points**



**Figure 9. Map with high density awareness regions shaded for one customer**

(ii) *The recommender system.* A main advantage of simulation is the ability to test real recommender systems. Our base case examines sales diversity under two systems, termed here  $r_1$  and  $r_2$ . In the taxonomy of Adomavicius and Tuzhilin (2005), both are memory-based, collaborative filters. Recommender  $r_1$  is the most basic collaborative filter: for a given user, it first finds the set  $N^*$  of the  $n$  most similar customers by using cosine similarity to compare vectors of purchase counts. It then recommends the most popular item among this group.<sup>7</sup> Formally, let  $sales$  be an  $I$  users  $\times$   $J$  items matrix of purchase counts, with  $sales_{ij}$  the  $(i,j)$  element,  $sales_i$  the row vector of user  $c_i$ 's purchases, and  $sales_{(j)}$  the column vector of product  $p_j$ 's sales. For a given user  $c_i$ , let

$$N^* := \operatorname{argmax}_N \sum_{c_j \in N} \cos(sales_i, sales_j) \quad \text{such that } |N| = n, i \neq j \quad (5)$$

The system then recommends product

<sup>6</sup> We have tested sensitivity to different numbers of consumers and products, higher dimensions, and other distributions (e.g., uniform, normal, and Pareto for each combination of consumers and products as well as normal mixture distributions to introduce the idea of segments). The specific Gini values vary, but the conclusions are qualitatively similar. The main sensitivity results are in the online appendix.

<sup>7</sup> An alternative is to use correlation (i.e. cosine on mean-centered data) to find the set of most similar users. We have tested both, and this does not qualitatively affect the results.

$$r_1: j^* = \arg \max_j \sum_{c_i \in N^*} sales_{ij} \quad (6)$$

Recommender  $r_2$  has one difference. When selecting the most popular product among similar users, candidate items are first discounted by their overall popularity in the entire population:

$$r_2: j^* = \arg \max_j \frac{1}{sales_{(j)}} \sum_{c_i \in N^*} sales_{ij} \quad (7)$$

The motivation for  $r_2$ 's popularity discounting is a belief that popular items are so obvious they should not be suggested. For example, if a consumer is expected to buy or be aware of a product with high probability, the firm should recommend something else if it wishes to generate incremental sales. Interviews with industry experts suggested such popularity discounting is common practice. Note,  $r_2$  is not the same as applying "term-frequency inverse-document frequency" weights (*tf-idf*) to algorithm  $r_1$ . *tf-idf* would insert discounting in the user similarity calculation (Breese et al. 1998), whereas  $r_2$  computes an undiscounted user similarity and discounts popularity in the final argmax of (7). In Section 7, we test other recommenders, including one with *tf-idf* weights, and show the results are directionally the same.

(iii) *Awareness*. Recommenders are valuable to consumers because they help overcome information asymmetry: the seller and other users may know of a product, but the given consumer may not. Recommenders share this information across the population. We assume each consumer is aware of a subset of the  $J$  products, and only items in his awareness set can be purchased. Once an item is recommended to a consumer, he is always aware of it in future periods. At the start, consumers are aware of many of the central products on the map plus a few items in their own neighborhood. These initial awareness states  $\{0,1\}$  for each consumer-product pair are sampled according to

$$P(c_i \text{ aware of } p_j) = \lambda e^{-distance_{0j}^2 / \theta} + (1 - \lambda) e^{-distance_{ij}^2 / \kappa \theta} \quad (8)$$

Above,  $distance_{0j}$  and  $distance_{ij}$  are respectively the Euclidean distances from the origin to product  $p_j$  and from consumer  $c_i$  to product  $p_j$ . The higher is  $\lambda$ , the more users are aware of central, mainstream products (left term), and the higher is  $1 - \lambda$  the more users are aware of products in their neighborhood.  $\theta$  and  $\kappa \theta$  are scaling parameters, determining how fast awareness decays with distance. Note that the users are not

aware of the same products: they are likely to overlap in their awareness of the central products but less so in the local products.

The awareness model for one consumer is shown visually in Figure 9 for  $\lambda=.75$ ,  $\theta=.35$ , and  $\kappa=1/3$ . We use these values for our base case. Setting  $\lambda=.75$  creates a market with consumers more aware of mainstream goods than niche ones. This assumption is consistent with a hit-oriented market or one in which mass advertising makes consumers aware of mainstream products. Under the opposite ( $\lambda < .5$ ), the base-case is already a market of niches and it only strengthens later results that diversity can decrease.  $\theta$  determines how many central products users know. Setting  $\theta=.35$  creates an easy to understand “radius 1” rule:  $e^{-1/.35} = .057 \approx 0$ . In other words, outside a radius of 1, the consumer is unlikely to be aware of the product. In our maps, about 40% of the products are within 1 unit from the origin; it is on this 40% of products that consumers are likely to overlap most in their awareness. The value  $\kappa$  determines awareness in the consumer’s own neighborhood. The value  $\kappa=1/3$  creates roughly a 0.5 radius rule. Outside the 0.5 radius, the consumer is unlikely to know about products, unless they are the central ones. The approach in selecting these parameters was to create an interpretable base case. In sensitivity analysis, we find the Gini can change for other parameter values but the results are directionally the same.<sup>8</sup>

(iv) *Choice model.* At each step of the simulation, a consumer either purchases an item in his awareness set or makes no purchase at all. We model this using the multinomial logit. The logit is well established in economics and marketing and has an axiomatic origin in random utility theory (for a Marketing application, see Guadagni & Little 1983). Consumer  $c_i$ ’s utility for product  $p_j$  at time  $t$  is defined as  $u_{ijt} := v_{ijt} + \varepsilon_{ijt}$ , where  $v_{ijt}$  is a deterministic component and  $\varepsilon_{ijt}$  is an i.i.d. random variable with extreme value distribution. Under these assumptions

$$P(c_i \text{ buys } p_j \text{ at } t \mid c_i \text{ aware of } p_j \text{ at } t) = \frac{e^{v_{ijt}}}{\sum_{k=1}^J e^{v_{ikt}}} \quad (9)$$

---

<sup>8</sup> If consumers know only the central products ( $\lambda=1$ ) the results are directionally the same. If consumers are aware of all products ( $\theta \rightarrow \infty$ ), the results are the same direction as well. The same holds if awareness is Pareto distributed instead of normal. These results are in the online appendix.

The unconditional probability is defined  $P(c_i \text{ buys } p_j \text{ at } t) = P(c_i \text{ aware of } p_j \text{ at } t)P(c_i \text{ buys } p_j \text{ at } t \mid c_i \text{ aware of } p_j \text{ at } t)$ . If a consumer is unaware of a product, the left term is zero, and he cannot buy it.

The deterministic component  $v_{ijt}$  is often modeled as a linear combination of a brand intercept, product attributes, and/or market-related covariates (e.g., price, promotion). In our context, since all relevant variables up to white noise are encompassed in the map of products and consumer points, we define the deterministic portion:

$$v_{ijt} := \text{similarity}_{ij} = -k \log \text{distance}_{ij} \quad (10)$$

where  $\text{distance}_{ij}$  is the Euclidean distance between consumer  $c_i$  and product  $p_j$ . Our choice of a log transformation from distance to similarity is consistent with prior research (e.g., Schweidel et al. 2007).<sup>9</sup>

The parameter  $k$  determines the consumer's sensitivity to distance on the map. The higher is  $k$ , the more the consumer only prefers the closest products. For our base case, as  $k$  ranges from 1 to 40, the Gini increases from .68 to .75. This range is consistent with several prior estimates of market concentration in media and e-commerce settings. An estimate for a major online clothing retailer is 0.70 (Brynjolfsson et al. 2007), an estimate for the music sales of debut albums is 0.724 (Hendricks & Sorensen 2007)<sup>10</sup>, and an estimate for the online book market is also near 0.75 (Chevalier & Goolsbee 2003)<sup>11</sup>. To fix a base case, we use  $k = 10$  because the 0.72 Gini it produces matches the average of the estimates above. This  $k$  forms our base case. For other values, the results change in magnitude but not direction.

Last, as noted, consumers may choose not to purchase at all. This is modeled by an outside good having the same distance to all users. The outside good approach is one common specification for modeling a no-purchase option (e.g., Chintagunta 2002). Our base cases uses a distance of .75 for this option, which implies the outside good's proximity is about 90<sup>th</sup> percentile (.87) for each consumer. That

---

<sup>9</sup> Other transformations have been used, and the literature does not have a single standard: for example,  $-k \cdot \text{distance}_{ij}$  in Elrod (1988);  $(\text{distance}_{ij})^k$  in DeSarbo and Wu (2001); and  $-k \cdot \log(\text{distance}_{ij})$  in Schweidel et al. (2007) with  $k$  a scaling parameter. While our base case uses the log transformation (e.g., Schweidel et al. (2007) and other references contained therein), we have tested sensitivity to the other specifications, and the results are not substantively different.

<sup>10</sup> The .724 could underestimate concentration because the authors' data excludes less successful artists. This may not affect their objective, which differs from that in this paper.

<sup>11</sup> The Zipf formulation can be equated to a power law, and from the power law a closed form expression for the Gini can be derived. A rank-on-sales coefficient of 1.17 in a power law implies a Gini of  $(2 \times 1.17 - 1)^{-1} = 0.75$ .

is, for each person, the outside good is closer than roughly 90% of the other goods. This means consumers have a fairly good outside option. If the outside good is farther, diversity in the base case increases: consumers substitute farther products for the outside good. The change in Gini under recommendations, however, is in the same direction.

(v) *Saliency*  $\delta$ . The term  $\delta$  is the amount by which a recommended product's saliency is temporarily increased in the consumer's choice set. The impact of the saliency boost is that the purchase probability for the recommended item  $j$  is the same as that for an item  $j'$  with  $v_{ij'} = v_{ij} + \delta$ . The functional form is analogous to the modeling of store displays in marketing (e.g., Guadagni & Little 1983), which might be considered an offline example of recommendations. The resulting choice probability is  $P(c_i \text{ buys } p_j \text{ at } t \mid c_i \text{ aware of } p_j \text{ at } t) = e^\delta e^{v_{ijt}} \left( \sum_{k \neq i} e^{v_{ikt}} + e^\delta e^{v_{ijt}} \right)^{-1}$ .

When  $\delta = 0$ , the recommender has a pure awareness effect. Recommended items enter the awareness set if not there already. When  $\delta > 0$ , the recommender also has a saliency effect, which increases the probability of buying the item (conditional on awareness). The saliency effect may exist for several reasons. First, consumers aware of many goods may have difficulty comparing all of them; recommended items become more salient in this comparison. Second, the saliency boost may reflect the ease of clicking a recommended item versus continuing to search through a firm's website. Last, saliency may capture persuasive effects. Recommendations often show an item's packaging and artwork, akin to a persuasive advertisement. We assume the combined effect is to increase saliency through  $\delta$ . Experiments have begun to demonstrate that recommendations can have influential effects beyond awareness (Senecal & Nantel 2004). This simultaneity of both effects, awareness and saliency, has parallels with advertising's informative and persuasive effects (e.g., Narayanan et al. 2005).

The saliency term  $\delta$  is a key parameter because it controls the strength of the recommender. For this reason, the paper's main results will be shown for a range of  $\delta$  and not a single point. To give some intuition for  $\delta$ , consider the purchase probability of the 75th percentile closest item on the map (with 50 products, this is the 13<sup>th</sup> closest item). In our normal maps, if  $\delta = 0$  the user chooses item 13 with  $<10^{-4} \approx$

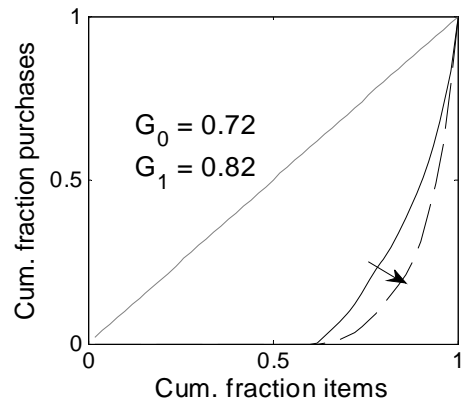
0 chance. Item 1 is purchased with probability 0.85. If the 75th percentile item is recommended, for  $\delta = (1, 5, 10, 15)$  the item takes on purchase probability ( $<10^{-3}$ ,  $<.01$ ,  $.15$ , and  $.48$ ) respectively. Thus  $\delta = 0-1$  is low, for it has little effect on purchase probability. A value  $\delta=15$  is rather high, for it makes a close item (100th percentile) and far item (75th percentile) equal in probability.

## 6. RESULTS

We now present simulation results for the two real-world recommenders. We use 50 consumer points and 50 products sampled from a bivariate normal  $N_2(\mathbf{0}, \mathbf{I})$  with  $k = 10$ . Each simulation is 50 iterations without recommendations and 50 iterations with recommendations.

### 6.1 Example of a Single Sample Path

Before presenting overall results, we illustrate the process with one sample run. At first, recommendations are disabled and customers make purchases for 200 periods. Then  $r_1$  is enabled and customers make purchases for an additional 200 periods. For sake of illustration,  $\delta = 5$ , but more general results follow. The Lorenz curves and Ginis from both periods are shown in Figure 10. The example shows  $G_1 - G_0 = 0.82 - 0.72 = 0.10 > 0$ , and hence  $r_1$  increases concentration here. This is for one sample path, and a more systematic comparison is given below.



**Figure 10. One sample path: before and after recommendations ( $r_1, \delta=5$ )**

### 6.2 Simulation Results

With the same parameters as above, we average results across 1000 experiments/maps each for  $r_1$  and  $r_2$ . After, we generalize the findings beyond the base case of  $\delta = 5$ .

As Table 1 shows, both recommenders have a concentration bias on average, as reflected by  $\bar{G}_1 > \bar{G}_2 > \bar{G}_0$  (0.81 > 0.74 > 0.72). The “standard” collaborative filter  $r_1$  has the larger bias. It is not surprising that  $\bar{G}_1 > \bar{G}_2$  because  $r_2$  explicitly discounts popularity. However, we do find it surprising that  $\bar{G}_2 > \bar{G}_0$ : beforehand, we could not rule out the possibility of  $r_2$ 's discounting leading to lower concentration. In fact, in a small number of runs (16%),  $r_2$  increases diversity, but in the majority of runs (84%) and on average it reduces diversity. A t-test of paired differences for unequal means (pre versus post recommendations) shows the differences are significant.

For  $r_1$ , this is partly explained by (6), in which popularity determines what product is recommended. This creates a self-reinforcing cycle: popular items are recommended more, items recommended more are purchased more, purchased items are recommended more, and so on. Despite this, the increased concentration was not readily obvious: recommendations are generated in many local user groups, making a priori conclusions difficult. While  $r_2$  dampens the popularity bias, the result also originates from using only sales data to make recommendations. Products with limited historical sales have little or no chance of being recommended even if they would be favorably received by the consumer.

Figure 11 shows the change in Gini for a range of  $\delta$ . When the recommender has both awareness and salience effects, concentration increases in  $\delta$ . The effect is most pronounced at high  $\delta$ , where by construction the recommender has a bigger effect. In the special case  $\delta = 0$ , the recommender has only awareness effects. System  $r_1$  continues to increase concentration, although by much less (+1.4%), as seen in Figure 11. The positive feedback loop is weakened: even if popular items are recommended more, recommended items are not necessarily purchased more because  $\delta = 0$ . As a result, the Gini's increase is attenuated. With  $r_2$ , diversity increases under  $\delta = 0$ , although the magnitude is small (-1.4%) as shown in Figure 11. The deliberate exploration of  $r_2$  coupled with low salience of recommendations increases diversity.

To summarize, when recommenders have both effects, diversity generally decreases. When recommenders affect only awareness, diversity decreases slightly for  $r_1$  and increases slightly for  $r_2$ . The  $\delta$

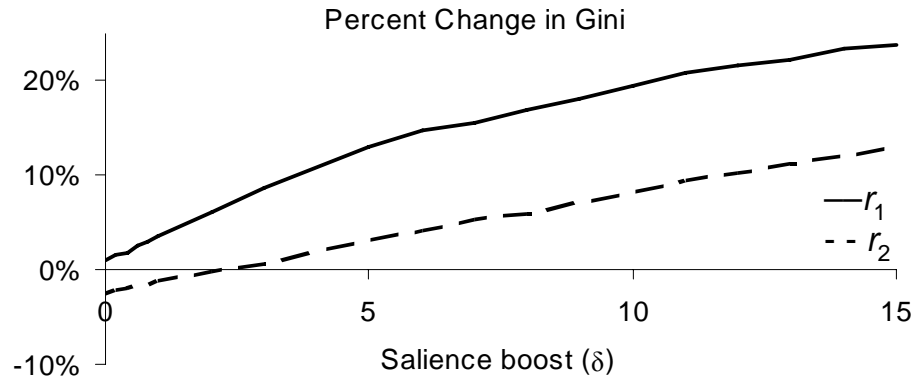


= 0 case is of conceptual interest, although it may not be commonplace if it is difficult to show consumers information without influencing them. As an example, the experiments of Senecal and Nantel (2004) show recommendations are influential even when consumers are aware of all products.

**Table 1. Comparison of  $r_1$  and  $r_2$  for  $\delta = 5$  (1000 experiments; parentheses give standard errors)**

	Average Gini		Average unique items aware of per person (AUIAP)		Average unique items bought per person (AUIBP)	
	$r_1$	$r_2$	$r_1$	$r_2$	$r_1$	$r_2$
Before	0.72 (.05)	0.72 (.05)	5.98 (1.33)	5.98 (1.33)	1.67 (0.32)	1.67 (0.32)
After	0.81 (.03)	0.74 (.04)	6.33 (1.34)	6.79 (1.41)	1.47 (0.17)	1.83 (0.26)
Change*	+0.09 (.03)	+0.02 (.02)	+0.37 (0.14)	+0.79 (0.24)	-0.20 (0.23)	+0.15 (0.25)

\* Change reports the average paired difference. T-test of paired differences is significant ( $<.05$ ) for the bottom row.



**Figure 11. Change in Gini by recommender and salience level ( $\delta$ )**

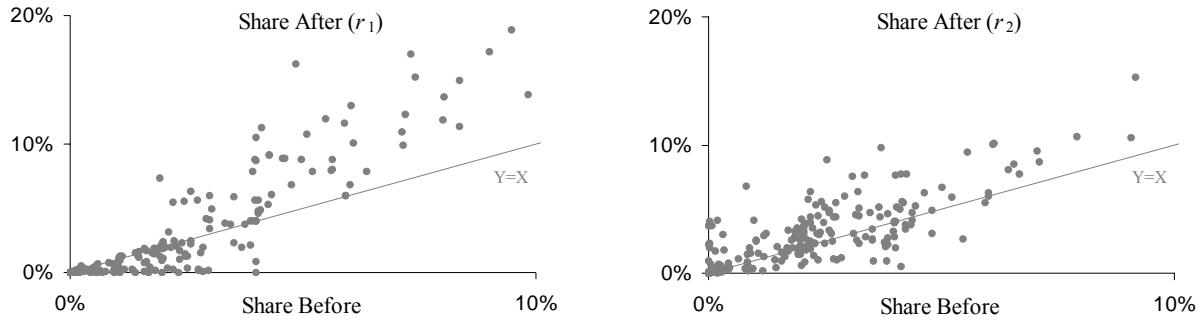
### 6.3 Further Discussion of the Results

This section contrasts aggregate versus individual-level diversity, product-level effects, and consumer-level effects. These three aspects give a better understanding of the model's mechanics and properties.

(i) *Aggregate versus individual effects.* Table 1's middle section shows the average unique items aware of per person (AUIAP). This quantity increases under recommendations. Systems  $r_2$ , as expected, creates a bigger increase, but in general both inform consumers of new products. Combining this observation with the change in Gini is revealing. Individually, consumers learn of more products (higher AUIAP), yet in aggregate diversity can decrease (higher Gini). This may explain users' anecdotes that recommenders create diversity even when an aggregate statistic, the Gini, reports more concentration.

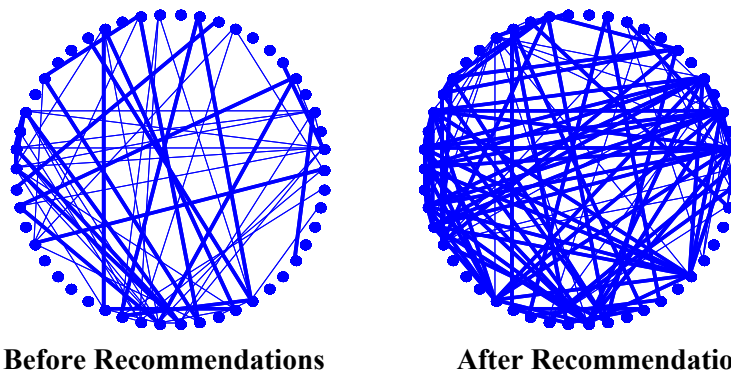
A similar effect is seen in Table 1's right panel. This panel reports the average unique items bought per person (AUIBP). Under  $r_1$ , consumers buy a narrower range of items, as seen by the lower AUIBP. Under  $r_2$ , the outcome is different: AUIBP increases. Consumers are pushed toward products that are not necessarily popular, which means they are less likely to have bought them previously. The Gini, however, still increases. This again leads to the finding that individual diversity can increase while aggregate diversity decreases. Consumers are discovering new products, but they are discovering the same products others have bought.

(ii) *Product-level view*. Figure 12 shows how the market share of particular products is affected by the recommender. Each point represents a product, with the  $x$  coordinate giving the product's market share before recommendations and the  $y$ -coordinate giving its share after. With recommendations, there is a systematic dispersion off the 45-degree line. The concentration bias is especially clear with  $r_1$ . Low share products become even lower, which is shown by the mass of points in the lower-left, and high share products become even higher, as seen by the mass of points in the upper-right. This reflects a 'poor get poorer' and 'rich get richer' phenomenon, both of which contribute to the increased Gini. The lower portion is related to the "cold-start" problem of collaborative filters, in which unpurchased/unrated items cannot be recommended (Schein et al. 2002). While the bias is not as acute with  $r_2$ , the high share products are likely to gain more share in this case as well. It is interesting to note that the positive feedback effect of recommendations can turn some medium selling products into high selling ones, which is consistent with the findings from the urn model in Section 4. As long as a product has modest sales, recommendations have the potential to make it more successful.



**Figure 12. Market shares by product ( $\delta=5$ ). Each point is a product whose coordinates give its market share before versus after recommendations. (Data pooled across 10 experiments)**

(iii) *Consumer-level view.* The recommender systems push consumers toward the same products, and thus make consumers more similar in their purchases. This is illustrated in Figure 13. In the graph, consumers are nodes equally spaced on the perimeter of a circle. An edge joins consumers  $(c_i, c_j)$  if  $\text{correlation}(\text{sales}_i, \text{sales}_j) > 0$ . The correlation is calculated over users' purchase counts for all items. Visually, the edge's thickness is proportional to the correlation. Comparing the graphs, the increased density at right shows that consumers have become more similar to one another in terms of products they consume. The figure alone does not imply the Gini has increased. For example, a correlation of 1 among all users could occur if everyone bought a single product (Gini=1), but it could also occur if all users bought all items equally (Gini=0). On its own, the figure shows consumers have become more alike. Combining this with the increased Gini, we see the complete picture: users are more similar (from Figure 13), and the items they purchase come from a smaller, more popular set (Ginis in Table 1).



**Figure 13. Consumer similarity ( $\delta=5, r_1$ ). Each point is a user. Users with common purchases are connected. Edge thickness is proportional to the number of items the two users have in common.**

## 7. SENSITIVITY ANALYSIS

We next examine our results under other possible assumptions. We approach this in four parts: alternate recommender systems; bestseller lists in the base case; variety seeking in the utility specification, and alternate parameter values.

### 7.1 Alternate Recommender Systems

The base case examined two recommenders that were considered representative of industry practice. This section tests additional systems. A comparison of eight recommenders  $r_i$  ( $i=1,\dots,8$ ) is given in Table 2. As before, we make comparisons of the form  $E[G_0] \approx \bar{G}_0$  (Gini without recommendations) versus  $E[G_i] \approx \bar{G}_i$  (Gini under  $r_i$ ).

The recommenders tested are as follows.  $r_1$  and  $r_2$  are as before.  $r_3$  is another popularity-discounting variation on  $r_1$  (Breese et al. 1998). It places discounting in the user similarity calculation but not the product selection calculation. (i.e.  $r_2$  and  $r_3$  add discounting in opposite places). Specifically, in (5) the user-item frequencies are multiplied by the inverse of each item’s total sales (known as the “inverse document frequency” (*idf*) in the field of information retrieval); once the similar user group is determined, the undiscounted argmax of (6) is used. This still leads to an increase in the Gini. The magnitude is similar to  $r_1$ ’s increase for the following reason. The intention of  $r_3$  is to prevent latently different users with little purchase history from being grouped together (e.g., two users who each bought Harry Potter and one very different item). Because of the initialization period, our users have several purchases, and so the similar user-groups under  $r_1$  and  $r_3$  are often similar (and hence  $\bar{G}_1 \approx \bar{G}_3$ ). System  $r_4$  is a combination of  $r_2$  and  $r_3$ : discounting is performed in both the user similarity calculation and argmax. As with its parents,  $r_4$  also lowers diversity.

To build context for these comparisons, we tested four other designs ( $r_5 - r_8$ ). System  $r_5$  recommends the lowest sales product. As expected, it decreases the Gini. System  $r_6$  recommends the median selling product. It also reduces the Gini because it diverts attention from otherwise higher selling products. System  $r_7$  recommends the best-selling product and as expected increases the Gini. We highlight that the

Gini under  $r_7$  is not higher than under  $r_1$ . A single product, the best seller, cannot be close to everyone. As a result, fewer users accept  $r_7$ 's recommendations, limiting its influence. In contrast,  $r_1$  recommends local best-sellers, which are closer to each user and thus accepted more.  $r_8$  is a best-seller list, which recommends the top 5 selling items. This system has the highest concentration: it shows the most popular items, and by showing multiple items increases the chance that at least one is close to the user and thus accepted. Similar results were confirmed experimentally by Salganik et al. (2006).

**Table 2. Comparison of additional recommenders (30 experiments).**

	$r_1$	$r_2$	$r_3$	$r_4$
	“Standard” CF	$r_1 + \text{argmax discount}$	$r_1 + \text{idf weights}$	$r_2$ and $r_3$ combined
$\bar{G}_i$	0.81 (0.03)	0.74 (0.05)	0.81 (0.03)	0.74 (0.05)
$\bar{G}_i - \bar{G}_0$ *	+0.09 (0.03)	+0.02 (0.02)	+0.09 (0.03)	+0.02 (0.02)
	$r_5$	$r_6$	$r_7$	$r_8$
	Lowest	Median	Best-Seller	Best-Seller List (Top 5)
$\bar{G}_i$	0.45 (0.10)	0.61 (0.03)	0.81 (0.04)	0.85 (0.03)
$\bar{G}_i - \bar{G}_0$ *	-0.27 (0.10)	-0.11 (0.04)	+0.09 (0.02)	+0.14 (0.04)

\* All results are significant at the 0.05 level (two-sided, paired differences t-test for unequal means). For all results, the no recommendations case has  $\bar{G}_0 = 0.72$  (0.05).

## 7.2 Bestsellers in the Base Case (Alternate Recommender Systems)

Without recommender systems, consumers might obtain product suggestions from best-seller lists. If the base case is driven by best-seller lists, any force that pulls consumers away from this influence would be expected to reduce concentration. We model this by introducing a best-seller list in the base case. This is equivalent to  $r_8$  from the previous subsection – only that  $r_8$  is now the control and  $r_1$  or  $r_2$  the treatment. Viewed this way, if recommenders are substituting for purely popularity-based filters, such as best-seller lists, they can increase diversity:  $\bar{G}_1 < \bar{G}_8$  ( $0.82 < 0.85$ ) and  $\bar{G}_2 < \bar{G}_8$  ( $0.75 < 0.85$ ). Although it is unlikely that best-seller lists drive purchase decisions in all product categories, it seems feasible that bestseller lists may be major drivers of purchase decisions in the absence of recommenders in at least some product categories. If so, this implies the role of recommenders is misunderstood. Relative to an ‘older’ world of bestseller lists, recommenders may reduce concentration, by virtue of cutting out the

even more popularity-biased tool ( $\bar{G}_1, \bar{G}_2 < \bar{G}_8$ ). But relative to a world without such lists, recommenders may increase concentration ( $\bar{G}_1, \bar{G}_2 > \bar{G}_0$ ).

### 7.3 Modifying the Utility Specification: Variety Seeking

Since the choice model allowed for repeat purchases, we ask whether the concentration results are affected if consumers seek variety across purchase occasions. The concept of state dependence has a long history in choice models (e.g., McAlister 1982). “Structural state dependence” (Seetharaman 2004) is the extent to which prior purchases of a product affect its future purchase propensity; positive dependence is termed inertia, while negative dependence is termed variety seeking.

To incorporate variety and inertia in the specification, we use a common approach and define

$$\begin{aligned} v_{ijt} &:= -k \log \text{distance}_{ij} + \beta X_{ijt} \\ X_{ijt} &:= \alpha X_{ijt-1} + (1 - \alpha) I(c_i \text{ bought } p_j \text{ on } t - 1) \end{aligned}$$

$X_{ijt}$  is an exponential smooth of purchase indicators  $I(\cdot)$ , and thus it summarizes how often and recently  $c_i$  has bought  $p_j$ . The parameter  $\alpha \in (0,1)$  determines how much weight is placed on recent versus distant purchase occasions.  $\beta$  determines the effect strength, with  $\beta < 0$  for variety seeking, and  $\beta > 0$  for inertia. This approach has been used frequently in the literature (e.g., see Guadagni & Little 1983; Seetharaman 2004). Past empirical studies have found consistent values of  $\alpha$  in the range 0.70-0.80, and thus we set  $\alpha = 0.75$  (Guadagni & Little 1983; Lattin 1987; Seetharaman 2004). For  $\beta$ , we consider a range of values to explore both variety-seeking and inertia.<sup>12</sup> The  $\beta$  term is not applied to the outside good, which by definition has the same distance to all consumers at all times.

Table 3 shows the Gini under state dependence. Under inertia ( $\beta > 0$ ), the findings are directionally the same as before: concentration increases. As  $\beta$  increases, the change in Gini dampens. Under high inertia, consumers do not want to deviate from their choices in the pre-recommendation period, and so the recommender’s influence becomes limited. Under variety seeking ( $\beta < 0$ ), concentration still increases for

---

<sup>12</sup> An alternative approach could replace  $X_{ijt}$  with a smooth on attribute levels of goods previously consumed (e.g., Lattin 1987). Our model uses an exponential smooth of purchase indicators, a common approach.

$r_1$  but by less.  $r_1$  suggests heavily purchased items, which are less likely to provide variety. As a result, users ignore recommendations that are too similar, and the change in Gini is lessened. For  $r_2$ , at moderate levels of variety seeking (e.g.,  $\beta = -5$ ) concentration still increases. At strong levels of variety seeking, the Gini can decrease. For example, at  $\beta = -20$ , the Gini drops .03 points. We note first that this level of variety seeking is high. Suppose  $c_i$  buys  $p_j$  semi-frequently so that  $X_{ijt} = 0.5$  at some time.  $\beta = -20$  implies  $\beta X_{ijt} = (-20)(0.5) = -10$ , which is twice as strong as the  $\delta = 5$  salience effect of recommendations. Under such high variety seeking, the Gini decreases because users ignore recommendations of popular items and selectively accept recommendations of less popular ones. Whereas  $r_1$  cannot supply these selectively accepted recommendations ( $r_1$  focuses on past hits),  $r_2$  makes this possible. Users want items not purchased recently, and  $r_2$ 's discounting better meets this goal.

The variety seeking results have an interesting interpretation. If consumers seek recommendations only in their most variety-seeking moments, diversity increases under  $r_2$ . However, as recommenders become ubiquitous, consumers are affected by them all the time, not only in one-off, high variety-seeking events. This may be the case for sites users visit regularly: for example, personalized news, personalized radio, and personalized retail websites. In these cases, increased concentration could be likely.

**Table 3. Gini values under state dependence at  $\delta=5$ . For variety seeking  $\beta < 0$ , and for inertia  $\beta > 0$ .**

$\beta$	-30	-20	-10	-5	0	5	10	20	30
$\overline{G_1 - G_0}$	+0.04	+0.05	+0.07	+0.08	+0.09	+0.07	+0.04	+0.02	+0.02
$\overline{G_2 - G_0}$	-0.04	-0.03	-0.01	+0.01	+0.02	+0.03	+0.02	+0.02	+0.02

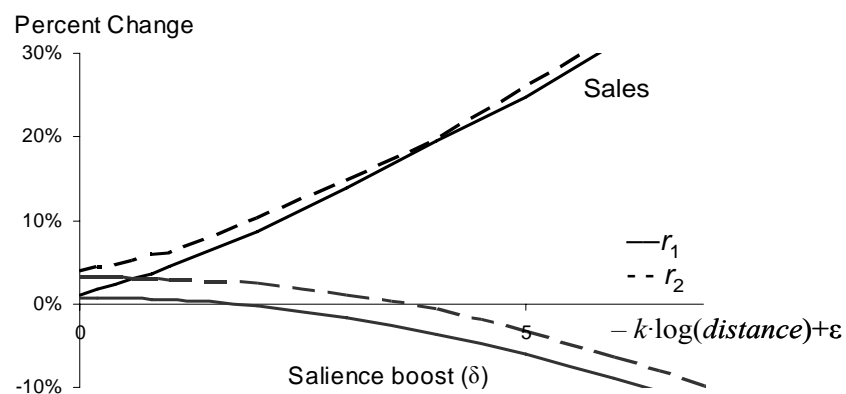
#### 7.4 Altering Other Simulation Parameters

We also examine sensitivity to other simulation parameters (e.g., number of consumers and products, map distributions). The main sensitivity results appear in the appendix, and others are available on request. In general, we find that varying these parameters affects the degree of the results (e.g., the Gini may increase more versus less), but the substantive findings remain the same.

## 8. WELFARE IMPLICATIONS

Thus far we have examined how recommenders affect concentration. We next ask whether these changes leave firms and consumers better off.

For firms, we examine the change in sales. For consumers, we examine the change in product fit. The consumers' product fit is defined as the average of  $-k \log \text{distance}_{ijt} + \varepsilon_{ijt}$  over all purchases, including those of the outside good. This quantity reflects the map distance between consumers and purchased products. Figure 14 shows these quantities. The figure puts the numbers in percent change so that both firm and consumer effects can be plotted on the same scale. When  $\delta = 0$ , the recommender has a pure awareness effect. The firm's sales are higher, and consumers find products closer to them. The gains for both parties are larger under  $r_2$ . The deliberate exploration of  $r_2$  helps consumers find better products, which translates into higher sales (fewer no-purchases) for the firm. When  $\delta > 0$ , recommenders have both awareness and salience effects. At low  $\delta$ , the results are the same as  $\delta = 0$ . At high  $\delta$ , firms always sell more: the greater the salience  $\delta$ , the more likely the consumer is to buy the recommended product than the outside good. For consumers, high  $\delta$  increases the average map distance of purchases. Consumers may forgo a slightly closer product if the recommended product has increased salience. A slightly better song or news article may be available deeper in the website, but the recommendation's salience makes it easier to click.



**Figure 14. Percent changes in consumer and producer surplus for varying levels of salience ( $\delta$ )**



Does this mean consumers are worse off when  $\delta > 0$ ? A myopic answer is yes. If the salience effect is simply a momentary increase in purchase probability but does not contribute to post-purchase satisfaction, then consumers are worse off because their purchases are farther away. A more complete answer considers additional factors. First, it is possible that  $\delta$ , or part of it, should be included in the consumer's utility. This is the case if recommendations add value to the choice occasion. In this case, the consumer effect in Figure 14 becomes positive and increasing (not shown for clarity). This view is consistent with several logit applications in marketing in which a store display adds utility to the choice occasion (e.g., Guadagni & Little 1983). For example, a display means the user does not have to walk down the aisle to get the product or price information. Similarly, choosing the recommended item may save time browsing the site or effort in making product comparisons. Second, to the extent media products have positive externalities, these may offset the increased distance. For example, watching the same movies as others is valuable because it permits discussion. In this case, the recommender serves a coordinating role whose value is not fully accounted for in  $-k \log distance_{ij}$ .

For firms, we have measured changes in sales. To the extent changes in concentration simplify inventory and supply chain management, these factors are also unaccounted for. Last, it should be noted that even if recommenders reduce consumer-product distances and increase sales, they may have welfare implications at the societal level. Legal scholar Cass Sunstein (2001) discusses the risk of "filters" creating a fragmented society. Sunstein asks whether en masse filtering of all but one's specific interests will reduce the ability of society's members to understand each other. Such considerations are beyond the current scope, but we raise them to show that an exhaustive analysis of welfare implications will involve more than changes in sales and map distances.

## **9. CONCLUSIONS AND FUTURE WORK**

This paper examined the effect of popular recommender designs on sales concentration and offered evidence that recommenders do influence sales diversity. In a range of cases, common recommenders were found to exert a concentration bias. Thus the traditional view that recommenders increase diversity

may not always hold. The work also demonstrated that some designs may be associated with greater bias than others. The results have important managerial and consumer implications. For retailers whose strategy is to offer variety, certain system designs may be at odds with that goal. We also find that recommenders can increase sales, and recommenders that discount popularity appropriately may increase sales more. For consumers, we showed that recommenders can create better (closer) consumer-product matches. However, if the recommendations are highly salient and thus influential, popularity-influenced recommendations may displace what would otherwise be better product suggestions. Future, empirical work would be valuable for determining the relative strength of these awareness versus salience effects.

Given these findings, why do consumers feel that recommendations have increased the range of media they consume? We offered several explanations. The first is that diversity can increase at the individual level but still decrease in aggregate. This was borne out under  $r_2$ , in which each user became aware of more items and purchased more unique items, but the Gini still increased. Individuals may be exploring more choices, but they are being pushed toward the same choices. Second, if recommenders are pulling consumers away from a world of best-seller lists, diversity can increase by virtue of cutting out the even more popularity-biased tool. A final possibility is that the effect of increased product offerings outweighs the effect of recommenders. Increased offerings may lower concentration (Anderson 2006; Brynjolfsson et al. 2007), while recommenders may temper but not reverse the effect. While it is beyond the current scope to examine the simultaneous effects of recommenders and increased offerings, this is an interesting question for future work.

A final interesting aspect arose to the extent that externalities exist for media goods. If, for example, there is a benefit to reading popular books or seeing popular movies (e.g., by increasing the likelihood of being able to discuss the experience with others), then consumer utility involves a tradeoff between a Hotelling-like similarity and the externality from a popular product. To the extent such externalities are strong, it would be interesting to see if they pose a limit, or upper bound, on the degree of diversity

consumers would ever prefer. If this were the case, a concentration bias may be more desirable than previously considered. We hope to explore these questions in future work as well.

## 10. REFERENCES

- Adomavicius, G. and A. Tuzhilin. 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17(6):734-749.
- Anderson, C. 2006. *The Long Tail*. New York: Hyperion.
- Ansari, A., S. Essegai, and R. Kohli. 2000. Internet recommendation systems. *J. of Marketing Research* 37(3):363-375.
- Arthur, W. B. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.
- Bikhchandani, S., D. Hirshleifer, and I. Welch. 1998. Learning from the behavior of others: conformity, fads, and informational cascades. *J. of Economic Perspectives* 12:3 151-170.
- Bodapati, A. 2008. Recommendation systems with purchase data. *J. Mktg. Res.* 45(1):77-93.
- Breese, J., D. Heckerman, and C. Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. of the 14<sup>th</sup> Conference of Uncertainty in AI*.
- Brynjolfsson, E., Y. Hu, and D. Simester. 2007. Goodbye Pareto principle, hello long tail: the effect of search costs on the concentration of product sales. *SSRN eLibrary* 953587.
- Brynjolfsson, E., Y. Hu, and M. Smith. 2006. From niches to riches: the anatomy of the long tail. *Sloan Management Review* 47(4) 67-71.
- Chevalier, J. and A. Goolsbee. 2003. Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics* 1(2):203-222.
- Chintagunta, P. 2002. Investigating category pricing behavior at a retail chain. *J. Mktg. Res.* 39(2):141-54
- Clemons, E.K., G. G. Gao, and L.M. Hitt. 2006. When online reviews meet hyperdifferentiation. *J. of Management Information Systems* 23(2):149-171.
- Cooke, A., H. Sujan, M. Sujan, and B. Weitz. 2002. Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations. *J. Mktg. Res.* 39(4):488-497.
- Dellarocas, C. 2003. The digitization of word-of-mouth: promise and challenges of online feedback systems. *Management Science* 49(10):1407-1424.
- DeSarbo, W. S. and J. Wu. 2001. The joint spatial representation of multiple variable batteries collected in marketing research. *Journal of Marketing Research* 38(2):244-253.
- Durrett, R. 2005. *Probability: Theory and Examples*. Belmont, CA: Thomson Brooks/Cole.

- Elrod, T. 1988. Inferring an ideal-point product-market map from consumer panel data. In *Data, Knowledge and Decisions*, p. 240-249. Edited by W. Gaul and M. Schader. Berlin: Springer-Verlag.
- Elrod, T. and M. P. Keane. 1995. A factor-analytic probit model for representing the market structure in panel data. *J. of Marketing Research* 32(1):1-16.
- Fleder, D. and K. Hosanagar. 2007. Recommender systems and their impact on sales diversity. In *Proc. of the 8<sup>th</sup> ACM Conference on Electronic Commerce*, p. 192-199.
- Guadagni, P.M. and J.D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* 2(3):203-238.
- Hendricks, K. and A. T. Sorensen. 2007. Information spillovers in the market for recorded music. *NBER working paper W12263*. Updated version downloaded 02/18/08 at <www.stanford.edu/~asorensen>.
- Hill, B. M., D. Lane, and W. Sudderth. 1980. A strong law for some generalized urn processes. *Annals of Probability* 8(2):214-226.
- Hotelling, H. 1929. Stability in competition. *The Economic Journal* 39:41-57.
- Johnson, N. L. and S. Kotz. 1977. *Urn Models and their Applications*. New York: John Wiley & Sons.
- Lattin, J. M. 1987. A model of balanced choice behavior. *Marketing Science* 6(1):48-65.
- Linden, G., B. Smith, and J. York. 2003. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7(1):76-80.
- McAlister, L. 1982. A dynamic attribute satiation model for choices made across time. *J. Cons. Res.* 9(3):141-150.
- Mooney, R. J. and L. Roy. 2000. Content-based book recommending using learning for text categorization. *5<sup>th</sup> ACM Conf. on Digital Libraries*, p.195-204.
- Narayanan, S., P. Manchanda, and P. K. Chintagunta. 2005. Temporal differences in the role of marketing communication in new product categories. *J Mktg. Res.* 42(3):278-290.
- Oestreicher-Singer, G. and A. Sundararajan. 2006. Network structure and the long tail of electronic commerce. Paper presented at The Wharton School, OPIM seminar. Philadelphia, PA. 9 January.
- Resnick, P. and H. Varian. 1997. Recommender systems. *Comm. of the ACM* 40(3):56-58.
- Salganik, M. J., P. S. Dodds, and D. J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854-856.
- Sarwar, B., G. Karypis, J. Konstan, and J. Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proc. of the 10<sup>th</sup> International World Wide Web Conference*.
- Schafer, J., J. Konstan, and J. Riedl. 1999. Recommender systems in e-commerce. *ACM EC*, p.158-166.

- Schein A. I., A. Popescul, L. H. Ungar, and D. M. Pennock. 2002. Methods and metrics for cold-start recommendations. *SIGIR*, p. 253-260.
- Schweidel, D. A., E. Bradlow, and P. Fader. 2007. Modeling the evolution of customers' service portfolios. *SSRN eLibrary* 985639
- Seetharaman, P. B. 2004. Modeling multiple sources of state dependence in random utility models: a distributed lag approach. *Marketing Science* 23(2):263-271.
- Sen, A. 1976. Poverty: an ordinal approach to measurement. *Econometrica* 44(2):219-231.
- Senecal, S. and J. Nantel. 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing* 80:159-169.
- Sunstein, C. R. 2001. *Republic.com*. Princeton: Princeton University Press.
- Van Alstyne, M. and E. Brynjolfsson. 2005. Global village or cyber-Balkans? Modeling and measuring the integration of electronic communities. *Management Science* 51(6):851-868.

## APPENDIX

This appendix proves all results stated for the analytic model. In the main paper, the propositions are stated in terms of segment market shares ( $X_t$ ). By noting that a share  $X_t$  is just a statistic of the products' sales, results can be derived in a random walks framework since the sales are integer quantities.

For ease of understanding the derivations, Propositions 2 and 3 are derived in reverse order. The numbering is consistent with the main paper, but the order of presentation is reversed.

Before deriving results specific to recommender systems, we need one preliminary result.

For a simple random walk on the integers, let

$$\begin{aligned} S &:= \text{Event \{particle at } i \text{ moves to } i+1 \text{ on next move\}} \\ \theta &:= P(S) \\ i \rightarrow j &:= \text{Event \{particle at } i \text{ ever reaches } j\}} \end{aligned}$$

**Lemma 1. (One-Away Return)**

$$\begin{aligned} P(i \rightarrow i+1) &= \begin{cases} \theta & , \theta < \frac{1}{2} \\ 1-\theta & , \theta \geq \frac{1}{2} \\ 1 & \end{cases} \\ P(i \rightarrow i-1) &= \begin{cases} 1-\theta & , \theta > \frac{1}{2} \\ \theta & , \theta \leq \frac{1}{2} \\ 1 & \end{cases} \end{aligned}$$

**Proof.**

The above is a basic result from stochastic processes (e.g., Durrett, 2005, p. 294).

**Recommendations context**

Turning to our context, for which we have products  $\{w, b\}$ , let

$$\begin{aligned} p &:= P(\text{consumer picks } w \text{ on own}) \\ r &:= P(\text{consumer follows recommendation}) \\ W_t, B_t &:= \text{Total } w, b \text{ in Urn 2 prior to purchase } t \\ Z_t &:= W_t - B_t \\ X_t &:= W_t/(W_t+B_t), \text{ which is } w\text{'s share before purchase } t \\ g(X_t) &:= P(w \text{ recommended at } t | X_t) \end{aligned}$$

Thus the chance a consumer selects  $w$  at  $t$  is

$$f(X_t) := P(\text{consumer buys } w \text{ at } t) = p(1-r) + g(X_t)r$$

As defined in the main paper,  $g$  is a step function. This implements a system that recommends the product with majority share.

$$g(X_t) := \begin{cases} 0 & , X_t < \frac{1}{2} \\ \frac{1}{2} & , X_t = \frac{1}{2} \\ 1 & , X_t > \frac{1}{2} \end{cases}$$

Substituting this  $g(X_t)$  into the expression  $f(X_t)$  gives

$$f(X_t) = \begin{cases} p(1-r) & , X_t < \frac{1}{2} & "l" \\ \frac{[p(1-r)] + [p(1-r) + r]}{2} & , X_t = \frac{1}{2} & "m" \\ p(1-r) + r & , X_t > \frac{1}{2} & "h" \end{cases}$$

The letters  $l, m, h$  are shorthand for the expressions at their left. They also have a geometric interpretation:  $f$  is a modified step function (shifted and stretched), and  $l, m,$  and  $h$  correspond to the heights of  $f$ 's lower segment, middle point, and upper segment. This graphical interpretation was shown in the main paper.

While the propositions state results about  $\{X_t\}$ , here we study  $\{Z_t\}$ . Studying sales instead of shares carries the same information because  $X_t$  is a statistic of sales. This switch, however, is beneficial because  $Z_t$  is an integer that changes by one unit each period, and can thus be studied in a random walks framework. (The reason for focusing on  $Z_t$  but not  $W_t$  and  $B_t$  will become clear below.)

For any time  $\tau$  at which  $Z_\tau = 0$  (i.e.  $W_\tau = B_\tau$ ), three events are possible

$$\begin{aligned} WB & := \text{Event } \{Z_t > 0 \text{ for all } t > \tau \mid Z_\tau = 0\} \\ BW & := \text{Event } \{Z_t < 0 \text{ for all } t > \tau \mid Z_\tau = 0\} \\ RTZ & := \text{Event } \{Z_t = 0 \text{ for some } t > \tau \mid Z_\tau = 0\} \end{aligned}$$

In words,  $WB$  is the event “ $w$  leads black forever after the next draw”;  $BW$  is the event “ $b$  leads  $w$  forever after the next draw”; and  $RTZ$  is the event “ $Z_t$  returns to zero at some future time point”.

By the definition of  $X_t$ ,  $Z_t > 0 \Rightarrow X_t > \frac{1}{2}$ ,  $Z_t < 0 \Rightarrow X_t < \frac{1}{2}$ , and  $Z_t = 0 \Rightarrow X_t = \frac{1}{2}$ . Thus the events  $WB, BW,$  and  $RTZ$  can also be interpreted as “white has majority share from the next draw on”, “black has majority share from the next draw on”, and “white and black will have equal shares at some future time point”. Thus  $\{Z_t\}$  carries information about  $\{X_t\}$ ; however,  $\{Z_t\}$ , unlike  $\{X_t\}$ , changes in increments of one each period, and is thus amenable to a random walks framework.

We now have a random walk on  $\mathbf{Z}^1$  beginning at the origin for which the transition probabilities are  $(l, 1-l), (m, 1-m),$  and  $(h, 1-h)$  depending on whether the particle is left of zero, at zero, or right of zero.

**Lemma 2. (Never Return Probabilities are always Non-Zero).**

For any  $p \in [0,1]$  and  $r \in (0,1)$ , either  $P(WB) > 0, P(BW) > 0,$  or both are  $> 0.$

Thus it is always possible for at least one product to obtain a majority share and never lose it.

**Proof.**

We prove this by giving probabilities for these events. By conditioning on the first event

$$\begin{aligned} P(WB) & = P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 \mid Z_{\tau+1} = 1) \\ P(BW) & = P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 \mid Z_{\tau+1} = -1) \\ P(RTZ) & = 1 - P(WB) - P(BW) \end{aligned}$$



Because the walk begins at the origin, the terms  $P(Z_{\tau+1} = 1)$  and  $P(Z_{\tau+1} = -1)$  follow immediately as  $m$  and  $1 - m$ . For the rightmost terms, Lemma 1 will be needed.

To apply the lemma, as will be seen, we must distinguish three cases. The interpretation of the three cases is given in the main paper. Further, we reparameterize the cases from  $(p, r)$  notation to  $(l, h)$  notation because it makes clearer how the lemma is applied.

The change of parameters assumes  $p \in [0, 1]$  and  $r \in (0, 1)$ , which is to say the recommender has some influence. The boundary case when  $r = 0$  or  $1$  is of less interest, for it does not concern recommender systems, but for completeness it will be discussed afterward.

Case 1.  $p \leq (\frac{1}{2} - r)/(1 - r) \Leftrightarrow l < \frac{1}{2}, h \leq \frac{1}{2}$ .

$$\begin{aligned} P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\ &= m \cdot [1 - P(1 \rightarrow 0)] \\ &= m \cdot (1 - 1) \\ &= 0 \end{aligned}$$

$$\begin{aligned} P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\ &= (1 - m) \cdot [1 - P(-1 \rightarrow 0)] \\ &= (1 - m) \cdot \left(1 - \frac{l}{1-l}\right) \end{aligned}$$

$$\begin{aligned} P(RTZ) &= 1 - P(WB) - P(BW) \\ &= 1 - (1 - m) \cdot \left(1 - \frac{l}{1-l}\right) \end{aligned}$$

Case 2.  $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r) \Leftrightarrow l < \frac{1}{2}, h > \frac{1}{2}$

$$\begin{aligned} P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\ &= m \cdot [1 - P(1 \rightarrow 0)] \\ &= m \cdot \left(1 - \frac{1-h}{h}\right) \end{aligned}$$

$$\begin{aligned} P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\ &= (1 - m) \cdot [1 - P(-1 \rightarrow 0)] \\ &= (1 - m) \cdot \left(1 - \frac{l}{1-l}\right) \end{aligned}$$

$$\begin{aligned} P(RTZ) &= 1 - P(WB) - P(BW) \\ &= 1 - m \cdot \left(1 - \frac{1-h}{h}\right) - (1 - m) \cdot \left(1 - \frac{l}{1-l}\right) \end{aligned}$$

Case 3.  $p \geq \frac{1}{2}/(1 - r)$  when  $\Leftrightarrow l \geq \frac{1}{2}, h > \frac{1}{2}$

$$\begin{aligned} P(WB) &= P(Z_{\tau+1} = 1)P(Z_t > 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = 1) \\ &= m \cdot [1 - P(1 \rightarrow 0)] \\ &= m \cdot \left(1 - \frac{1-h}{h}\right) \end{aligned}$$

$$\begin{aligned}
P(BW) &= P(Z_{\tau+1} = -1)P(Z_t < 0 \text{ for } t > \tau + 1 | Z_{\tau+1} = -1) \\
&= (1-m) \cdot [1 - P(-1 \rightarrow 0)] \\
&= (1-m) \cdot (1-1) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
P(RTZ) &= 1 - P(WB) - P(BW) \\
&= 1 - m \cdot \left(1 - \frac{1-h}{h}\right)
\end{aligned}$$

Reviewing the above expressions shows that for every case either  $P(WB) > 0$ ,  $P(BW) > 0$ , or both are  $> 0$ .

□

Recall that the parameter space is the unit square  $\{(p,r): 0 \leq p,r \leq 1\}$ . The above three cases are non-overlapping and cover the space  $\{(p,r): 0 \leq p \leq 1 \wedge 0 < r < 1\}$ . To be exhaustive and cover the entire square, we point out the two remaining trivial cases. When  $r = 0$ , this gives a Bernoulli process that converges to  $p$  by the weak law of large numbers. Setting  $r = 1$  forces all consumers to purchase whatever was bought on the first occasion, which is  $w$  with the chance of a fair coin flip. The resulting behaviors are clear, but the model does not apply to recommender systems unless  $r \in (0,1)$ .

The above lemma showed it is always possible for a product to obtain a majority share and never lose it. Next we show this must be the case: after sufficient time, one product is guaranteed to develop a majority share and maintain. Further, how likely it is for  $w$  versus  $b$  to obtain such a position is also given next.

**Lemma 3. (Probability of  $w$  versus  $b$  obtaining a lasting majority share).**

Case	$\lim_{t \rightarrow \infty} P(Z_t > 0)$	$\lim_{t \rightarrow \infty} P(Z_t < 0)$	$\lim_{t \rightarrow \infty} P(RTZ)$	Interpretation
1	0	1	0	$b$ wins always
2	$1 - \gamma$	$\gamma$	0	either can win
3	1	0	0	$w$ wins always

where  $\gamma = \frac{(1-m) \cdot \left(1 - \frac{l}{1-l}\right)}{m \cdot \left(1 - \frac{1-h}{h}\right) + (1-m) \cdot \left(1 - \frac{l}{1-l}\right)}$

**Proof.**

$$\begin{aligned}
& \lim_{t \rightarrow \infty} P(Z_t > 0) \\
&= \sum_{i=1}^{\infty} P(WB \text{ occurs after } i^{\text{th}} \text{ time } w = b) \\
&= \sum_{i=1}^{\infty} P(RTZ)^{i-1} P(WB | RTZ \text{ occurs } i-1 \text{ times}) \\
&= \sum_{i=1}^{\infty} P(RTZ)^{i-1} P(WB) \\
&= P(WB) \sum_{i=0}^{\infty} P(RTZ)^i \\
&= \frac{P(WB)}{1 - P(RTZ)} \\
&= \frac{P(WB)}{1 - [1 - P(WB) - P(BW)]} \\
&= \frac{P(WB)}{P(WB) + P(BW)}
\end{aligned}$$

The analogous argument gives

$$\lim_{t \rightarrow \infty} P(Z_t < 0) = \frac{P(BW)}{P(WB) + P(BW)}.$$

We can also confirm that

$$\begin{aligned}
& \lim_{t \rightarrow \infty} P(RTZ) \\
&= \lim_{t \rightarrow \infty} \{1 - P(Z_t > 0) - P(Z_t < 0)\} \\
&= 1 - \frac{P(WB)}{P(WB) + P(BW)} - \frac{P(BW)}{P(WB) + P(BW)} \\
&= 0
\end{aligned}$$

Combining the above expressions with the results from the previous lemma gives

Case	$\lim_{t \rightarrow \infty} P(Z_t > 0)$	$\lim_{t \rightarrow \infty} P(Z_t < 0)$
1	$\frac{0}{0 + (1-m) \cdot \left(1 - \frac{l}{1-l}\right)} = 0$	$\frac{(1-m) \cdot \left(1 - \frac{l}{1-l}\right)}{0 + (1-m) \cdot \left(1 - \frac{l}{1-l}\right)} = 1$
2	$\frac{m \cdot \left(1 - \frac{1-h}{h}\right)}{m \cdot \left(1 - \frac{1-h}{h}\right) + (1-m) \cdot \left(1 - \frac{l}{1-l}\right)} = 1 - \gamma$	$\frac{(1-m) \cdot \left(1 - \frac{l}{1-l}\right)}{m \cdot \left(1 - \frac{1-h}{h}\right) + (1-m) \cdot \left(1 - \frac{l}{1-l}\right)} = \gamma$
3	$\frac{m \cdot \left(1 - \frac{1-h}{h}\right)}{m \cdot \left(1 - \frac{1-h}{h}\right) + 0} = 1$	$\frac{0}{m \cdot \left(1 - \frac{1-h}{h}\right) + 0} = 0$

□

The above result shows that in the limit (i) some product must develop and maintain a majority share and (ii) how often  $w$  versus  $b$  has this lasting majority. We now determine what those limiting shares are.

**Proposition 1. (Support of  $\lim_{t \rightarrow \infty} X_t$ ).** As  $t \rightarrow \infty$ ,  $\{X_t\}$  converges to either one or two values given by

Case	$\lim_{t \rightarrow \infty} X_t$	
	Outcome 1	Outcome 2
1	$l$	n/a
2	$l$	$h$
3	$h$	n/a

**Proof.**

Case 1.

By Lemma 3,  $\lim_{t \rightarrow \infty} P(Z_t < 0) = 1$ .

Thus  $\lim_{t \rightarrow \infty} X_t < 1/2$  because  $Z_t < 0 \Rightarrow X_t < 1/2$  by the definition of  $X_t$

Because  $\lim_{t \rightarrow \infty} X_t < 1/2$  and  $f$  is constant for  $X_t < 1/2$ ,  
 $\lim_{t \rightarrow \infty} X_t = f(X_t | X_t < 1/2) = l$ .

In words, because  $\lim_{t \rightarrow \infty} P(Z_t < 0) = 1$  this implies  $w$  always has the minority share. If  $w$  has the minority share, then  $\lim_{t \rightarrow \infty} X_t$  must be  $< 1/2$  by definition. Now  $f(X_t) = l$  everywhere on  $[0, 1/2)$ . Thus after sufficient time,  $\lim_{t \rightarrow \infty} P(\text{Draw } w) = f(\tau) = l$  for  $\tau < 1/2$ .

Said differently, if  $w$  has the minority share, its chance of being purchased is  $l$ ; since it was proved to remain the minority share product for this case, its limiting share must also be  $l$ .

Case 2.

By the previous result,  $\lim_{t \rightarrow \infty} P(Z_t > 0) \in (0, 1)$ ,  $\lim_{t \rightarrow \infty} P(Z_t < 0) \in (0, 1)$ , and  $\lim_{t \rightarrow \infty} P(RTZ) = 0$ . Thus either  $\lim_{t \rightarrow \infty} Z_t > 0$  or  $\lim_{t \rightarrow \infty} Z_t < 0$ .

Suppose we are in a path for which  $\lim_{t \rightarrow \infty} Z_t < 0$ . By the argument for Case 1 above,  $\lim_{t \rightarrow \infty} \{X_t | Z_t < 0\} = l$

In contrast, suppose we are in a path for which  $\lim_{t \rightarrow \infty} Z_t > 0$ . By the argument for Case 3 below,  $\lim_{t \rightarrow \infty} \{X_t | Z_t > 0\} = h$ .

Since one of these outcomes is guaranteed to occur (some product will have a non-reversing majority share), then  $\lim_{t \rightarrow \infty} X_t \rightarrow X$  where  $X$  is a random variable with support  $\{l, h\}$ .

Case 3.

By Lemma 3,  $\lim_{t \rightarrow \infty} P(Z_t > 0) = 1$ .

Thus  $\lim_{t \rightarrow \infty} X_t > \frac{1}{2}$  because  $Z_t > 0 \Rightarrow X_t > \frac{1}{2}$  by the definition of  $X_t$

Because  $\lim_{t \rightarrow \infty} X_t > \frac{1}{2}$  and  $f$  is constant for  $X_t > \frac{1}{2}$ ,  
 $\lim_{t \rightarrow \infty} X_t = f(X_t | X_t > \frac{1}{2}) = h$

The interpretation is analogous to Case 1 above.

□

**Corollary 1.** In Case 2,  $P(\lim_{t \rightarrow \infty} X_t < \frac{1}{2}) > 0$  and  $P(\lim_{t \rightarrow \infty} X_t > \frac{1}{2}) > 0$ .

Thus regardless of the initial preference  $p$ , either product can obtain and maintain a majority share.

**Proof.**

By Proposition 1,  $\{X_t\} \rightarrow X$  where  $X$  is a random variable with support  $\{l, h\}$ . Since  $l < \frac{1}{2}$  and  $h > \frac{1}{2}$ , as verified next, the result is shown.

Regarding  $l$ , Case 2 defines that  $p < \frac{1}{2(1-r)}$ . Thus  $l \equiv p(1-r) < \frac{(1-r)}{2(1-r)} = \frac{1}{2}$ .

Regarding  $h$ , Case 2 defines that  $p > \frac{(\frac{1}{2}-r)}{(1-r)}$ . Thus  $h \equiv p(1-r) + r > \frac{(\frac{1}{2}-r)(1-r)}{(1-r)} + r = \frac{1}{2}$ .

□

**Proposition 3. Distribution of  $\lim_{t \rightarrow \infty} X_t$ .** The distribution of  $\lim_{t \rightarrow \infty} X_t$  is

Case	Support point 1	Support point 2	$P(\lim_{t \rightarrow \infty} X_t = \text{support point 1})$	$P(\lim_{t \rightarrow \infty} X_t = \text{support point 2})$
1	$l$		1	0
2	$l$	$h$	$\gamma$	$1 - \gamma$
3	$h$		1	0

where  $\gamma = \frac{(1-m) \cdot (1 - \frac{l}{1-l})}{m \cdot (1 - \frac{1-h}{h}) + (1-m) \cdot (1 - \frac{l}{1-l})} \in (0,1)$ .

**Proof.**

The support points were derived in Proposition 1. The probabilities are determined as follows. For Cases 1 and 3, we know from Proposition 1 there is convergence to a single outcome; thus chance of converging to that particular outcome is 1. For Case 2, by Proposition 1 the process converges to one of two limiting outcomes,  $l$  and  $h$ . By Corollary 1, outcome  $l$  means  $b$  has majority share and outcome  $h$  means  $w$  has majority share. By Lemma 3,  $b$  versus  $w$  obtains the majority share with chance  $\gamma$  versus  $1 - \gamma$ , which means that  $l$  and  $h$  also occur with probabilities  $\gamma$  and  $1 - \gamma$ .

□

Thus far, the lemmas and propositions have been used to derive the distribution of  $\lim_{t \rightarrow \infty} X_t$ . With the limiting behavior of  $\{X_t\}$  understood, we now ask whether that limiting behavior reflects more or less concentration.

**Proposition 2. Relation of limit points to concentration.**

The term “increased concentration” refers to shares that are less equal than they would be without recommendations. Formally, “increased concentration” means  $\lim_{t \rightarrow \infty} X_t > p$  when  $p > \frac{1}{2}$  and  $\lim_{t \rightarrow \infty} X_t < p$  when  $p < \frac{1}{2}$ . When  $p = \frac{1}{2}$ , increased concentration occurs when  $\lim_{t \rightarrow \infty} X_t \neq p$ .

As before, we have  $p \in [0, 1]$  and  $r \in (0, 1)$ . (The case of  $r = 0$  or  $1$  was discussed above). The effect on concentration is then given by the following:

Case	Support points	Effect on concentration relative to $p$
1. $p \leq (\frac{1}{2} - r)/(1 - r)$	1	Increased concentration
2. $(\frac{1}{2} - r)/(1 - r) < p < \frac{1}{2}/(1 - r)$	2	Case 2A: $p \in (\frac{1-r}{2-r}, \frac{1}{2-r})$ . Increased concentration for both support points Case 2B: $p \notin (\frac{1-r}{2-r}, \frac{1}{2-r})$ . Increased concentration for one support point; decreased for the other
3. $p \geq \frac{1}{2}/(1 - r)$	1	Increased concentration

**Proof.**

Case 1.

The process converges to  $l \equiv p(1 - r) < p$ . Since  $p < \frac{1}{2}$  in Case 1, this implies increased concentration.

To verify that  $p < \frac{1}{2}$ , start with Case 1’s definition  $p \leq (\frac{1}{2} - r)/(1 - r)$ . Viewing  $p$  as a function of  $r$ , its derivative is  $dp/dr = -(1 - r)^{-1} + (\frac{1}{2} - r)(1 - r)^{-2}$ . The condition  $r \in (0, 1)$  implies  $dp/dr < 0$  on  $(0, 1)$ , and thus  $p(r)$  is maximized as  $r \rightarrow 0$ . Since  $p(0) = (\frac{1}{2} - 0)/(1 - 0) = \frac{1}{2}$ , this bound shows  $p$  must be less than  $\frac{1}{2}$  on the interval  $(0, 1)$ .

### Case 2.

First consider the case when  $p < 0.5$ . The two possible limits are  $p(1 - r)$  and  $p(1 - r) + r$ . Note that  $p(1 - r) < p$ , and thus this outcome always involves increased concentration. Now, consider the other outcome  $p(1 - r) + r$ . Since Case 2's definition states  $p > (\frac{1}{2} - r)/(1 - r)$ , it follows that  $p(1 - r) + r > \frac{1}{2}$ . Clearly, this reverses the popularity order of the two products. However, it increases concentration only if  $p(1 - r) + r > (1 - p)$ . Simplifying this expression, concentration increases only if  $p > (1 - r)/(2 - r)$ . Similarly, for the case in which  $p > 0.5$ , concentration increases in both outcomes only if  $p < 1/(2 - r)$ . Combining results, we see that sales concentration always increases if  $p \in (\frac{1-r}{2-r}, \frac{1}{2-r})$ . Otherwise, sales concentration increases for one limiting outcome and decreases for another.

### Case 3.

The process converges to  $h \equiv p(1 - r) + r = p + r(1 - p) > p$ . Since  $p > \frac{1}{2}$  in Case 3, this implies increased concentration.

To verify that  $p > \frac{1}{2}$ , start with Case 3's definition  $p \geq \frac{1}{2}/(1 - r)$ . Viewing  $p$  as a function of  $r$ , its derivative is  $dp/dr = \frac{1}{2}(1 - r)^{-2}$ . Adding the condition  $r \in (0, 1)$  implies that  $dp/dr > 0$  on  $(0, 1)$ , and thus  $p(r)$  is minimized as  $r \rightarrow 0$ . Since  $p(0) = \frac{1}{2}/(1 - 0) = \frac{1}{2}$ , this bound shows  $p > \frac{1}{2}$  on the interval  $(0, 1)$ .

□

## References

Durrett, R. 2005. *Probability: Theory and Examples*. Belmont, CA: Thomson Brooks/Cole.