Harbingers of Failure

Harbingers of Failure

Eric Anderson, Song Lin, Duncan Simester and Catherine Tucker

January 2015

We show that some customers, whom we call 'Harbingers' of failure, systematically purchase new products that flop. Their early adoption of a new product is a strong signal that a product will fail - the more they buy, the less likely the product will succeed. Firms can identify these customers either through past purchases of new products that failed, or through past purchases of existing products that few other customers purchase. We discuss how these insights can be readily incorporated into the new product development process. Our findings challenge the conventional wisdom that positive customer feedback is always a signal of future success.

Keywords: new product development, early adopters, lead users, preference heterogeneity

Introduction

Decades of research emphasizes that customer feedback is a critical input throughout the new product development process. A central premise of this customer-focused process is that positive feedback is good news. The more excited that customers are about a prototype, the more likely it is that a firm will continue to invest. When firms move to the final stages of testing and launching a new product, metrics of success shift from likes and dislikes to actual product sales. Here again, conventional wisdom is that more product sales indicate a greater likelihood of long-term success. This assumption is fundamental to nearly every new product forecasting model (Bass 1969; Mahajan, Muller and Bass 1990).

In this paper, we challenge this commonly held assumption. We show that not all positive feedback should be viewed as a signal of future success. In particular, using detailed transaction data from a chain of convenience stores, we show that increased sales of a new product by some customers can be a strong signal of future failure.

At first glance, the result is striking. How can more product sales signal future failure? After all, meeting sales targets is the litmus test of all new products. We present evidence that this result is driven by the existence of an unrepresentative subset of customers. We label these latter customers 'Harbingers' of failure. Harbingers are more likely to purchase products that other customers do not buy, and so a purchase by these customers may indicate that the product appeals to a narrower slice of the marketplace. This yields a signal that the product is more likely to fail.

We identify these customers in two ways. Our primary focus is on customers who have previously purchased new products that failed. We show that the tendency to buy hits or flops is systematic. If a customer tends to buy failures, then the next new product they purchase is more likely to be a failure. For example, customers who tend to purchase a successful product like a Swiffer mop are more likely to buy other ultimately successful products, like Arizona Iced Tea. In contrast, if a customer tends to buy flops, then their next purchase is also more likely to fail. For example, customers who purchased Diet Crystal Pepsi are more likely to have purchased Frito Lay Lemonade (both of which failed).

It is not only the initial purchase of new products by Harbingers that is informative, but also the decision to purchase again. A one-time purchase of Diet Crystal Pepsi is partially informative about a consumer's preferences. However, a consumer who repeatedly purchases Diet Crystal Pepsi is even more likely to have unusual preferences, and is more likely than other customers to choose other new products that will fail in the future.

The second way to identify Harbingers focuses on purchases of existing products. This second approach is motivated by evidence that customers who systematically buy new products that fail are also more likely to buy niche existing products. This suggests that a tendency to purchase niche existing products may identify Harbingers. The findings reveal that both approaches are similarly effective at identifying Harbingers, and that distinguishing between early adopters of new products using either metric can significantly improve predictions of long term success or failure.

Related Literature

Our results complement several literatures, including literatures on preference minorities, representativeness, lead users, and new product forecasting. We discuss each of these next.

We identify Harbingers either through past purchases of new products that failed, or through past purchases of existing products that few other customers buy. In both cases, Harbingers reveal preferences that are unusual compared to the rest of the population. The term "preference minorities" was previously coined by Joel Waldfogel (and his co-authors) to describe customers with unusual preferences. The existence of these customers has been used recently to explain the growth of Internet sales in some product categories. Offline retailers tend to allocate their scarce shelf space to the dominant preferences in that market, so customers whose preferences are not representative may not find products that suit their needs (Anderson 1979; Waldfogel 2003). Choi and Bell (2011) show that, as a result, preference minorities are more likely to purchase from the Internet, and are less price sensitive when doing so (see also Brynjolfsson, Hu and Rahman 2009). Preference minorities also help to explain why we see a longer tail of niche items purchased in Internet channels, compared to other retail channels (Brynjolfsson, Hu and Smith 2003; Brynjolfsson, Hu and Simester 2011).[1]

Lack of representativeness of customer preferences also underpins Geoffrey Moore's (1991) explanation that new technology products often fail because they are unable to "cross the chasm." He argues that early adopters of technology are more likely to be technology enthusiasts and visionaries and argues that the mainstream market has different (more risk averse) preferences. Early success may therefore not be a predictor of future success. We caution that Moore's explanation is focused primarily on the adoption of disruptive new technologies that represent significant innovations over existing products. The role of technology enthusiasts is less apparent in the consumer packaged goods markets that we study.

Van den Bulte and Joshi (2007) formalize Moore's explanation by modeling the diffusion of innovation in markets with segments of "influentials" and "imitators." They show that diffusion in such a market can exhibit a dip between the early and later parts of the diffusion curve, but that this depends on the extent to which the influential segment affects the imitators. They offer five different theories of consumer behavior that may explain this result. When extended to our setting, these theories may partially explain why we observe Harbingers making systematically different purchasing decisions than other customers, and why the new products that they purchase tend to fail. However, many of these theories also predict that one segment of customers will influence the purchasing decisions of other customers. In contrast, our explanation does not require that a group of customers influence the decisions of others. Moreover, in our consumer packaged goods setting, it is not obvious

[1] Huang, Singh and Srinivasan (2014) use a similar explanation to argue why some crowdsourcing participants may offer worse ideas than other participants.

that dependency between different customers' purchasing decisions is as large an effect as it is in technology markets.

This is not the first paper to recognize that the feedback of certain customers should be weighted differently in the new product development process. In particular, the lead user literature argues for giving *greater* weight to positive feedback from some customers. Rather than relying on information from a random or representative set of customers, the lead user process proposes collecting information from customers on the "leading edges" of the market. The rationale for this approach is that these leading customers are more likely to identify "breakthrough" ideas that will result in product differentiation (von Hippel 1986). There have been many studies seeking to validate these benefits. For example, Urban and von Hippel (1988) study the computer-aided design (CAD) market and show that reliance on lead users results in new product concepts that are preferred by potential users over concepts generated by traditional product development methods. Lilien, Morrison, Searls, Sonnack and von Hippel (2002) report findings from a natural experiment at 3M, exploiting variation in the adoption of lead user practices across 3M's business units. They find that annual sales of product ideas generated using a lead user approach are expected to yield eight times more revenue than products generated using traditional approaches. Our results complement this literature; while early adoption by lead users may presciently signal new product success, there also exist customers whose adoption is an early signal of product failure.

All of the new product introductions we study had survived initial pilot testing. Yet despite these screens, only 40% of the new products in our data survive for three years. This raises the question: How did the products that failed ever make it through the initial market tests? The new product development literature has identified escalated commitment (Boulding, Morgan and Staelin 1997; Brockner and Rubin 1985; Brockner 1992,), an inability to integrate information (Biyalagorsky, Boulding and Staelin 2006) and distortions in management incentives (Simester and Zhang 2010) as possible explanations. Our identification of this class of Harbingers provides an alternative explanation. If customers who initially adopt the product have unusual preferences that are different from other customers, then high initial sales may not signal future success.

Most new product forecasting models focus on predicting new product outcomes using an initial window of sales. Perhaps the best-known new product forecasting model was introduced by Bass (1969). An important characteristic of the Bass model is an assumed interaction between current adopters and potential adopters of the new product. The speed of diffusion depends on the degree to which later adopters imitate the early adopters. As we discussed, our explanation does not require dependency between different customers' purchasing decisions (and this effect may be relatively weak in the types of markets that we study). Moreover, a central prediction of these models is that a positive initial response is a signal of positive future outcomes.[2] Our findings indicate that this central premise may not hold if the positive initial response reflects purchases by Harbingers.

Another stream of new product forecasting models focuses on predicting success *before* the product has been launched on the market. The absence of market data means that pre-market forecasts are generally considered less accurate than forecasts that use an initial window of post-launch sales. However, if the cost of launch is sufficiently high, then pre-market tests can provide information to evaluate whether to invest in a new product launch. Urban and Hauser (1993) review nine different approaches to pre-market testing. Some approaches rely on experience with past new products to estimate the relationship between advertising, promotion and distribution response functions (see for example the NEWS model proposed by Pringle, Wilson and Brody 1982). Other approaches obtain estimates of trial and repeat purchase rates, and use these as inputs in a dynamic stochastic model to estimate cumulative sales (see for example Eskin and Malec 1976). The estimates of the trial and repeat parameters are obtained from various sources. For example, some models use pre-market field tests (Parfitt and Collins 1968), while others use laboratory tests. Perhaps the best known of the laboratory models is ASSESSOR, proposed by Silk and Urban (1978).

[2] In a recent study, Morvinski, Amir and Muller (2014) show that this prediction holds only when the initial positive response is from similar customers and when the uncertainty about the product quality is low.

Respondents are exposed to advertising and given an opportunity to purchase in a simulated store. The laboratory purchase rates are combined with estimates of availability and awareness to predict initial market adoption, while repeat purchase rates are estimated from mail order repurchases.[3]

The principle underlying models of trial and repeat purchasing is that adoption can be influenced by the firm's investments in advertising, distribution and promotion. However, long-term success depends upon customers accepting the product, often to the exclusion of a product they were previously using (Parfitt and Collins 1968; Eskin 1973; Fader and Hardie 2005). Repeat purchase rates may therefore provide a more accurate predictor of new product success than initial adoption rates. For this reason, we use both initial adoption and repeat purchases to classify customers. Specifically, we ask whether customers who repeatedly purchase new products that fail provide a more accurate signal of new product failure than customers who only purchase the new product once.

The remainder of the paper is organized as follows. In the next section, we describe the data in detail, including a summary of how long unsuccessful new products remain in the store and the opportunity cost of these failures to the retailer. Then, we present initial evidence that there exist customers whose decision to adopt a new product is a signal that the product will fail. We also conduct a wide range of checks to evaluate the robustness of the findings. Next, we investigate who the Harbingers are, and explore whether they can also be identified through purchases of (existing) niche items. We summarize our findings and their implications in the last section.

Data and Initial Results

This paper uses two datasets: a sample of individual customer transaction data, and a sample of aggregate store-level transaction data. Both datasets come from a large chain of

[3] A final class of premarket forecasting models compares customers' attitudes to the new product with attitudes to existing products. The challenge for these attitude approaches is to accurately map customer attitudes to actual purchase probabilities.

convenience stores with many branches across the United States. The store sells products in the beauty, consumer healthcare, edibles and general merchandise categories. Customers visit the store frequently (on average almost weekly), and purchase approximately four items per trip at an average price of approximately $4 per item.

The store level transaction data includes aggregate weekly transactions for every item in a sample of 111 stores spread across 14 different states in the Midwestern and Southwestern portions of the US. The data period extends from January 2003 through October 2009. We use the store-level transaction data to define new product survival, and to construct product covariates for our multivariate analysis. We exclude seasonal products that are designed to have a short shelf life, such as Christmas decorations and Valentine's Day candy.

The individual-customer data covers over ten million transactions made using the retailer's frequent shopping card between November 2003 and November 2005 for a sample of 127,925 customers. The customers represent a random sample of all of the customers who used the frequent shopping card in the 111 stores during this period. Their purchase histories are complete, and record every transaction in any of the firm's stores (in any geographic region) using the retailer's frequent shopping card. We focus on purchases of new products that were made between November 2003 and November 2005 and are within 52 weeks of the product's introduction. There are 77,744 customers with new product purchases during this period. They purchase a total of 8,809 different new products with a total of 439,546 transactions distributed across 608 product categories. Examples of these new products include: Paul Mitchell Sculpting Foam Mousse, Hershey's wooden pencils, SpongeBob children's shoelaces, and Snackwell sugar free shortbread cookies.

New Product Success

We will initially define a product as a "failure" if its last transaction date (in the store-level transaction data) is less than 3 years after its introduction.[4] If the last transaction date is after

[4] Product success is evaluated using store-level data, which contains purchases by all customers and includes customers who purchased without using a card. To the extent that customers using a loyalty card are different

this date then the product is a "success". This definition of success is relatively strict (the product has to survive a minimum of 36 months), and so we also investigate how robust our findings are to using a shorter survival horizon. In addition, we consider several alternative measures of success, including accuracy in a holdout sample, the market share of the new item, and how long the item survived in the market.

Across the full sample of 8,809 new products, 3,508 (40%) survived for 3 years (12 quarters).[5] Other research has reported similar success rates for new products in consumer packaged goods markets. For example, Liutec, Du and Blair (2012) cite success rates of between 10% and 30%. Similarly, Barbier et al. (2005), report success rates of between 14% and 47%. In evaluating whether a success rate of 40% is high or low, it is important to recognize that these new products have all survived the retailer's initial market tests and are now broadly introduced across the retailer's stores. If we were able to observe the full sample of new products that were either proposed by manufacturers or that were subjected to initial market tests by this retailer, then the success rate would be considerably lower.

The Cost of New Product Failure

As a preliminary investigation, we compared profits earned in the first year of a new product's life for new products that ultimately succeeded or failed. Throughout the first year, flops contribute markedly lower profits than hits do. This has an important implication for the retailer. Because shelf space is scarce (the capacity constraint is binding) the retailer incurs an opportunity cost when introducing a new product that fails (and when keeping them in the stores). Introducing a flop and keeping it for one year results in lost profits equivalent to 49% of the average annual profits for an existing item in the category. The implication is that new

---

from other customers, we would expect this to make it more difficult to predict store-level outcomes. Therefore, any selection bias introduced by the loyalty cards hinders rather than contributes to our findings. Similarly, we will use individual customer transactions at all of the firm's stores, not just the 111 stores for which we have aggregate weekly data (recall that all of the customers not only made purchases in the 111 stores during our data period, but they also made purchases in other stores). We will later investigate how this affects the results by limiting attention to purchases in the 111 stores, or by only considering purchases outside those stores.

[5] The average survival duration for the 5,301 failed items was 84.68 weeks, or approximately 21 months. In the Appendix we report a histogram (Figure A1) describing how long new products survive.

product mistakes are very costly. The more accurately the retailer can predict which new products will succeed (and the faster it can discontinue flops), the more profitable it will be.

It is important to recognize that throughout this paper we do not make any distinctions based on the reasons that a new product fails. Instead we will consider which customers purchase new products that succeed or fail, and identify customers ('Harbingers') whose purchases signal that a new product is likely to fail. We will also show that the retailer can more quickly identify flops if it distinguishes *which customers* initially purchase a new product, rather than just *how many customers* purchase. We begin this investigation in the next section.

Do Purchases by Some Customers Predict Product Failure?

Firms often rely on customer input to make decisions about whether to continue to invest in new products. Our analysis investigates whether the way that firms treat this information should vary for different customers. In particular, we consider the retailer's decision to continue selling a new product after observing a window of initial purchases, and show how this decision can be improved if the retailer distinguishes between Harbingers and non-Harbingers. Although our initial analysis focuses on customers prior purchases of new products that failed, we later also investigate whether we can identify Harbingers through their prior purchases of existing products. In particular, we identify customers who tend to purchase niche items that few other customers purchase.

Our initial analysis proceeds in two steps. First, we use a sample of new products to group customers according to how many flops they purchased in the weeks after the product is introduced. We then investigate whether purchases in the first 15 weeks by each group of customers can predict the success of a second sample of new products. We label this 15 weeks the "initial evaluation period." We demonstrate the robustness of the findings by varying how we select the groups of products, varying the length of the initial evaluation period that we use to predict new product success, and varying the metrics used to measure success.

The unit of analysis is a (new) product, and we assign the products into two groups according to the date of the new product introduction. We will initially assign products introduced between November 2003 and July 2004 to product set A (the "classification" set). The classification set contains 5,037 new products, of which 1,958 (38.9%) survive 3 years. New products introduced between July 2004 and July 2005 are assigned to product set B (the "prediction" set).[6] The prediction set contains 2,935 new products, including 1,240 (42.2%) successes. We will later investigate varying these demarcation dates and also randomly assigning products to the prediction and classification sets.

If the classification and prediction sets contain new products that are variants of the same item, this may introduce a spurious correlation between the failure rates for the two groups of items. For example, it is possible that the classification set includes a new strawberry flavored yogurt and the prediction set includes a new raspberry flavor of the same yogurt. It is plausible that the success of these products is correlated as the firm may choose to continue or discontinue the entire product range. For this reason, we restrict attention to new products for which there is only a single color or flavor variant. This ensures that products in the validation set are not merely a different color or flavor variant of a product in the classification set. It also ensures that the products are all truly new, not just new variants of an existing product.

Grouping Customers Using the Classification Product Set

To group customers according to their purchases of products in the classification set, we calculate the proportion of new product failures that customers purchased in the initial evaluation period and label this as the customer's $FlopAffinity_i$:

$$FlopAffinity_i = \frac{\text{Total number of flops purchased from classification set}}{\text{Total number of new products purchased from classification set}}. \quad (1)$$

[6] Transactions in the first 15 weeks after a new product is introduced are used to predict the product's success. Therefore, we cannot include products introduced between July 2005 and November 2005 in the prediction set as we do not observe a full 15 weeks of transactions for these items.

We classify customers who have purchased at least two new products during the product's first year into four groups, according to their *FlopAffinity*. These include 29,436 customers, which represents approximately 38% of all the customers in the sample. The 25th, 50th, and 75th percentiles are a flop rate of 25%, 50%, and 67%. Therefore, we use the following four groupings:

Group 1: Between 0% and 25% flops (25th percentile) in the classification set

Group 2: Between 25% and 50% flops (50th percentile) in the classification set

Group 3: Between 50% and 67% flops (75th percentile) in the classification set

Group 4: Over 67% flops in the classification set

Although we use these percentiles to group customers, the number of customers in each group varies because a disproportionally large number of customers have a flop rate of 0%, 50% or 100%. The groups sizes (for groups 1 through 4) are: 8,151 (28%), 4,692 (16%), 10,105 (34%) and 6,515 (22%). There are also 48,308 "other" customers who are not classified into a group, because they did not purchase at least two new products in the classification set.

The focus of the paper is investigating whether these different groups of customers can help predict success or failure of products in the prediction set.

Predicting the Success of New Products in the Prediction Set

Recall that the prediction set includes new products introduced after the products in the classification set. We will use purchases in the initial evaluation period (the first 15 weeks) to predict the success of the new products in the prediction set. We estimate two competing models. The competing models are both binary logits, where the unit of analysis is a new product indexed by j and the dependent variable, $Success_j$, is a binary variable indicating whether the new product survived for at least 3 years. The first model treats all customers equally, while the second model distinguishes between the four groups of customers. In particular, the probability that product *j* is a success (p_j) is modeled as:

Model 1: $$\ln\left(\frac{p_j}{1-p_j}\right) = \alpha + \beta_0 Total\ Sales_j \quad (2)$$

Model 2: $$\ln\left(\frac{p_j}{1-p_j}\right) = \alpha + \beta_1\ Group\ 1\ Sales_j + \beta_2 Group\ 2\ Sales_j + \beta_3 Group\ 3\ Sales_j + \beta_4 Group\ 4\ Sales_j + \beta_5 Sales\ to\ Other\ Customers_j \quad (3)$$

The *Total Sales* measure counts the total number of purchases of the new product during the initial evaluation period. The *Group x Sales* measures count the total number of purchases by customers in Groups 1-4. The *Sales to Other Customers* measures purchases by customers who are not classified (because they did not purchase at least two new products in the classification set). Therefore, the difference between the two models is that the first model aggregates all sales without distinguishing between customers. The second model distinguishes purchases according to the outcomes of the customers' classification set purchases.

These sales measures are all calculated using purchases in the first 15 weeks after the new product is introduced. Average marginal effects from both models are reported in Table 1. We report the likelihood ratio comparing the fit of Models 1 and 2, together with a Chi^2 test statistic measuring whether the improvement in fit is significant. We also report the area under the ROC curve (AUC). The AUC measure is used in the machine learning literature, and is equal to the probability that the model ranks a randomly chosen positive outcome higher than a randomly chosen negative outcome.[7] The closer this is to one, the more accurate the classifier. For both of these metrics (and for the additional success metrics we report later in this section), we compare Models 1 and 2. This represents a direct evaluation of whether we can boost predictive performance over standard models that forecast the

[7] The ROC curve represents the fraction of true positives out of the total actual positives vs. the fraction of false positives out of the total actual negatives.

outcome of new products using the number of initial purchases. In later analysis we will extend this comparison to not only consider initial trial of the product, but also repeat purchases (see for example Eskin 1973).

Table 1 about here

The Chi-square statistics and AUC measures both confirm that distinguishing initial purchases by Harbingers from other customers can significantly improve decisions about which new products to continue selling. Recall that the dependent variable is a binary variable indicating whether the product succeeded. Positive marginal effects indicate a higher probability of success, while negative marginal effects indicate the reverse. In Model 1 we see that higher *Total Sales* is associated with a higher probability of success. This is exactly what we would expect: products that sell more are more likely to be retained by the retailer. In Model 2 we observe positive marginal effects for customers in Groups 1 and 2, but negative marginal effects for customers in Groups 3 and 4. Purchases by customers in each group are informative, but they send different signals. Notably, purchases by Harbingers (represented by customers in Groups 3 and 4) are a signal of failure: If sales to these customers are high, the product is more likely to fail.

Our two models do not contain any controls for product covariates. However, we can easily add covariates to each model. To identify covariates we focus on variables that have previously been used in the new product development literature (see for example Henard and Szymanski 2001). Our focus is on controlling for these variables rather than providing a causal interpretation of the relationships. A detailed definition of all the control variables is provided in the Appendix, together with summary statistics (Tables A1 and A2). The findings when including product covariates are also included in Table 1 (as Models 3 and 4). We see that that the negative effect of sales from Group 4 on new product success persists.

Are Repeat Purchases More Informative?

Previous research has suggested that it is not just a customer's initial purchase of a product that is informative, but also whether the customer returns and repeats that purchase (Eskin 1973; Parfitt and Collins 1968). Therefore, we investigate whether repeat purchases of a new product that subsequently failed are more informative about which customers are Harbingers than just a single purchase. To address this question, we identify new products that are purchased repeatedly by the same customer in the classification set (during the initial evaluation period). In particular, we redefine *FlopAffinity* as:

$$\textit{Repeated FlopAffinity}_i(n) = \frac{x_i(n)}{y_i(n)} \quad (4)$$

where $x_i(n)$ equals the number of flops customer *i* purchased at least *n* times in the classification set, and $y_i(n)$ equals the total number of new products customer *i* purchased at least *n* times in the classification set.

We use this definition to group customers by their *Repeated FlopAffinity*. To investigate the relationship between the *Repeated FlopAffinity* groups and the success of products in the prediction set we vary the minimum number of purchases, *n*, from one to three. There are fewer customers who make repeat purchases of the products in the classification set, so we aggregate customers in Groups 1 and 2 and customers in Group 3 and 4. The findings are reported in Table 2.

Insert Table 2 about here.

When *n* is equal to 1, the definition in Equation (4) is equivalent to that in Equation (1). We include it to facilitate comparison. Using the new definition, we find the same pattern that sales from Harbingers significantly reduce the likelihood of success. Notably, the marginal effects for Groups 3 and 4 are larger as *n* becomes larger. This suggests that a new product is even more likely to fail if the sales come from customers who repeatedly purchase flops.

Embracing the Information that Harbingers Provide

Notice that the findings indicate that the firm should not simply ignore purchases by Harbingers, because their purchases are informative about which products are likely to fail. In particular, if we omit purchases by customers in Groups 3 and/or 4, the model is less accurate in explaining which products succeed. Notice also that when sales to Groups 3 and 4 groups are included, the model is able to give greater weight to purchases by other customers whose adoption is a strong signal that the product will succeed. This is reflected in the much larger positive marginal effect of sales to customers in Groups 1 and 2 (Model 2) compared to the marginal effect for *Total Sales* (Model 1). The large, significant effect for Groups 1 and 2 (see also Table 2, Model 2 At Least 1 purchase) provides some evidence that there may also exist customers whose purchases signal that a product is more likely to succeed (i.e., Harbingers of Success).

A simple counterfactual exercise can illustrate the danger of simply ignoring purchases by Harbingers. Suppose that half of the demand is contributed by customers who have high *FlopAffinity* (Group 4), while the remaining sales are equally distributed between the other three groups (and the unclassified "other" customers). Using the calibrated parameters, we predict that the probability of success drops from 39.51% to 0.67% as total unit sales increase from 0 to 100. By contrast, if half of the demand comes from Group 1 (and the other half is equally distributed across the other three groups and the "other" customers), then the probability of success increases from 39.51% to 71.83% as sales increase from 0 to 100 units. These results are illustrated in Figure 1. Distinguishing between these customers leads to very different predictions of product success.

Figure 1 about here

We can also illustrate how the probability of success changes as the fraction of sales contributed by Harbingers increases. We merge Groups 3 and 4 and define these customers as Harbingers (customers with *FlopAffinity* between 0.5 and 1). We calibrate the models with the merged groups (as reported in Table 2) and report the findings in Figure 2. Holding *Total Sales* fixed (at the average level), as the percentage of sales contributed by Harbingers increases from 25% to 50%, the probability of success decreases by approximately 31%. The success probability decreases even faster when we group customers using repeat purchases (*Repeated FlopAffinity*). Using at least two purchases to group customers, the probability of success drops 37%, and when using at least three purchases the drop is 56% (as the fraction of sales contributed by repeat Harbingers increases from 25% to 50%).

Figure 2 about here

Summary

We have presented evidence that customers who have tended in the past to purchase new products that failed can help signal whether other new products will fail. The signal is even stronger if these customers purchase the new product repeatedly. In the next section we investigate the robustness of this result.

Robustness Checks and Results by Product Category

In this section we conduct several checks to evaluate the robustness of the findings, including (a) alternative measures of product success, (b) alternative approaches of constructing product sets in the analysis, (c) alternative predictors of success. We also assess predictive accuracy using an alternative approach to construct the data. Finally, we report the findings by category and when grouping the items by other product characteristics. The findings for all of the results in this section are briefly summarized here, with detailed findings reported in the Appendix. In some cases, details of the analysis are also relegated to the Appendix.

Alternative Measures of Product Success

In the analysis reported in the previous section we focused on the same measure of product success: whether the product survived or not after three years. However, there are other measures of product success that we could consider. First, we replicate the analysis using a two-year survival window to define product success. The pattern of findings is unchanged under this new definition (see Table A3).

The next measure of success that we investigate is market share. Recall that the products in the prediction set were all introduced between July 2004 and July 2005. To measure market share, we calculate each item in the prediction set's share of total category sales in calendar year 2008 (approximately 3 years later). The qualitative conclusions do not change (see Table A4). Purchases by customers in Group 4 are associated with a significantly lower market share, and both the Chi-square and the AUC measures confirm that distinguishing between customers in Model 2 yields a significant improvement in accuracy.

An alternative to measuring whether a new product survives for two or three years is to measure how long a new product survives. One difficulty in doing so is that some products survive beyond the data window and so we do not have a well-defined measure of how long these products survive. To address this challenge, we estimate a hazard function. Details of the analysis are reported in the Appendix (Tables A5a and A5b). The findings reveal the same pattern as our earlier results; increased sales among customers in *FlopAffinity* Groups 3 and 4 are associated with higher hazards of product failure. The implication is that increased purchases by Harbingers are an indication that a new product will fail faster.

Alternative Constructions of Product Sets

Recall that when predicting the success of new products in the prediction product set, we only consider purchases made within 15 weeks of the new product introduction. We repeated the analysis when using initial evaluation period lengths of 5 or 10 weeks (using the same sample of products). The findings are again qualitatively unchanged (see Table A6). Even as early as five weeks after a new product is introduced, purchases by Harbingers are a significant predictor of new product success.

In our analysis, we omit items that are discontinued during the 15 week initial evaluation period. If we are using the initial evaluation period to predict future success, it seems inappropriate to include items that have already failed. However, these are not a random sample of items; they are the items that failed the most quickly. To investigate how the omission of these items affected the results, we repeated the analysis when including these items in our sample of new products. Including these items yields the same pattern of findings (see Table A7).

We have grouped products according to the timing with which they were introduced. New products purchased in the first 39 weeks of the data period (between November 2003 and July 2004) were assigned to the classification set, while products introduced between July 2004 and July 2005 were assigned to the prediction set. We repeated the analysis when using different time periods to allocate products into these two product sets. In particular, we constructed both a smaller classification set (the first 26 weeks) and a larger classification set (the first 52 weeks). The results confirm that the conclusions are robust to varying the length of the periods used to divide the products (see Table A8).

We also investigated two alternative approaches to constructing these two product sets. In one approach, we randomly assign the new products into the classification set and the prediction set, instead of dividing them by time. The classification set is again used to group customers while the prediction set is used to predict product success. Our qualitative conclusions remain unchanged under this approach. To further confirm that the classification set and prediction set contain new products that are truly different and not just related to each other, we also repeated the analysis when randomly assigning all of the new products in some product categories to the classification set and all of the new products in the remaining product categories to the prediction set. Product categories were equally likely to be assigned to each set. The findings also survive under this allocation (see Table A9).

Recall that our transaction data includes the complete purchasing histories of each customer in the sample (when using the store's loyalty card). This includes purchases from other stores in the chain, beyond the 111 stores used to construct covariates and identify when a product is introduced and how long it survived. To investigate how the findings are affected by the inclusion of purchases from other stores, we repeated the analysis using three different approaches. First, we excluded any purchases from outside the 111 stores when either classifying the customers into *FlopAffinity* groups, or when predicting the success of new products in the prediction set. Second, we only considered purchases from outside the 111 stores when classifying customers and predicting new product success. Finally, we obtained a sample of detailed transaction data for different customers located in a completely different geographic area, and used purchases by these customers to both classify these customers and predict new product success.[8] In the first two approaches, we used our original allocation of products to the classification and prediction sets. In the final approach, we randomly assigned new products into the classification set and the prediction set. As we might expect, the findings are strongest when we focus solely on purchases in the 111 stores, and weakest when we use customers from a different geographic area. However, even in this latter analysis, increased initial sales to Harbingers (customers in Groups 3 and 4) are associated with a lower probability of success (see Table A10).

Alternative Predictors of Success

We restricted attention to customers who have purchased at least two new products in the classification set. For customers with two purchases *FlopAffinity* can only take on values of 0, 0.5 or 1. To investigate whether the findings are distorted by the presence of customers who purchased relatively few new products, we replicated the analysis when restricting attention to customers who purchased at least three new products, at least four new products and at least five new products from the classification set. There is almost no qualitative difference in the results when restricting attention to a subset of customers using a minimum number of new product purchases (see Table A11). We conclude that the results do not

[8] This data includes 27 million transactions between August 2004 and August 2006 for a sample of 810,514 customers. These customers are a random sample of all of the customers who shopped in 18 stores located in a different geographic region. Their purchase histories are also complete, and record every transaction in any of the firm's stores (in any geographic region). Only 0.03% of the store visits for this separate sample of 810,514 customers were made at the 111 stores that we use in the rest of our analysis.

appear to be distorted by the presence of customers who purchased relatively few new products.

The variables in Equation (3) focus on the quantity purchased by customers in each *FlopAffinity* group. Alternatively, we can investigate how the probability of success varies when we focus on the *proportion* of purchases by customers in each group. In particular, we estimate the following modification to Equation (3):

$$\ln\left(\frac{p_j}{1-p_j}\right) = \alpha + \beta_o Total\ Sales_j + \beta_1 Group\ 2\ Ratio_j + \beta_2 Group\ 3\ Ratio_j + \beta_3 Group\ 4\ Ratio_j + \beta_4 No\ Sales\ to\ Grouped\ Customers \quad (5)$$

The *Ratio* measures represent the percentage of sales of product *j* to customers in Groups 1 through 4 that are contributed by customers in each group. The four ratio measures add to one (by definition), so we omit the *Group 1 Ratio* measure from the model. Under this specification the coefficient for each of the other three ratio measures can be interpreted as the change in the probability of success when there is an increase in the ratio of sales to that group (and a corresponding decrease in the ratio of sales to customers in Group 1). As we discussed, some items have no sales to any grouped customers. The ratio measures are all set to zero for these items and we include a binary indicator flagging these items. Table A12 in the Appendix reports the results. For Groups 3 and 4 we observe significant negative coefficients. They indicate that when a higher proportion of sales are contributed by customers in these groups, there is a smaller probability that the new product will succeed. In particular, if the ratio of sales contributed by customers in Group 3 increases by 10%, then the probability of success drops by 1.73%. A 10% increase in the Group 4 ratio leads to a 3% drop in the probability of success.

In our analysis we looked at purchases of new products by each group of customers. We can also investigate whether the decision not to purchase a new product is informative. In particular, we can classify customers according to the number of successful new products in the classification set that they did not purchase in the first year the product is introduced. We calculate the following measure:

$$\textit{Success Avoidance}_i = \frac{\text{Number of successful new products } NOT \text{ purchased by customer } i}{\text{Total number of new products } NOT \text{ purchased by customer } i}. \quad (6)$$

As we might expect, the *Success Avoidance* and *FlopAffinity* measures are highly correlated (r = 0.76). We use the same approach to investigate whether distinguishing between customers with high and low *Success Avoidance* can help to predict whether new products in the prediction set will succeed. The findings confirm that purchases by customers who tend to avoid success are also indicative of product failure (see Table A13).

Predictive Accuracy

An alternative way to measure predictive accuracy is to fit the models to a subset of the data and predict the outcomes in the remaining data. We divided the 2,953 products in the prediction set into an estimation sample and a holdout sample.[9] We used the estimation sample to produce coefficients and then use these coefficients to predict the outcome of the products in the holdout sample. The estimates of marginal effects are very close to those from the full sample and are reported in the Appendix (Table A14).

A baseline prediction would be simply that all of the items in the holdout sample will fail. This baseline prediction is correct 54.08% of the time. Using *Total Sales* during the 15-week initial evaluation period (Model 1) only slightly improves accuracy to 54.99%. However, in Model 2, where we distinguish which customers made those purchases during the initial evaluation period, accuracy improves to 61.42%.

This result highlights the value of knowing who is purchasing the product during an initial trial period. While higher total sales are an indication of success, this is only a relatively weak signal. The signal is significantly strengthened if the firm can detect how many of

[9] We use new items that are introduced earlier for estimation (60%) and holding out new items that are introduced later for prediction test (40%). The sample sizes are 1,740 in the estimation sample and 1,213 in the holdout sample. The findings are robust to randomly assigning the items to the estimation sample and the holdout sample.

those initial purchases are made by Harbingers. It is also useful to recognize that while a 6.43% improvement in accuracy (from 54.99% to 61.42%) may appear small, the difference is significant and meaningful. As we discussed in the literature review, the cost to the retailer of retaining bad products that will subsequently fail is large, in some cases even larger than the cost to the manufacturer. As a result, even small improvements in accuracy are valuable.

A limitation of this holdout analysis is that the outcomes (success or failure) for products in the estimation sample are not known when the items in the holdout sample are introduced (i.e. at the time of the prediction). Given we require three years of survival to observe the outcome, and we only have two years of individual purchasing data, there is no way to completely address this limitation. However, we offer two comments. First, in practice firms have access to longer data periods, and will be able to observe the outcomes of items in the calibration sample using data that exists at the time of the predictions. Second, it is important to distinguish information about product outcomes from information about customer's purchasing decisions. While the predictions use future information about the product outcomes, they only rely on customer purchases made prior to the date of the predictions.

Results by Product Category

We next compare how the findings varied across product categories. We begin by comparing the results across four different "super-categories" of products: beauty products, edibles, general merchandise and healthcare products. These super-categories are defined by the retailer,[10] and comprise 49%, 5%, 20% and 26% (respectively) of the new products in our sample of 2,953 new products in the prediction set. We repeated our earlier models separately for the four different super-categories (using the new products in the prediction set). The findings are reported in the Appendix (Table A15).

We replicate the negative effect of sales for Group 4 customers in each category, and the result is statistically significant in three of four categories. Further, distinguishing purchases

[10] These categories are considerably broader than the product categories used to allocate products in our earlier robustness analysis.

by the four customer groups appears to lead to particularly large improvements in accuracy in the healthcare and edibles categories. It could be argued that these are the categories where evaluating quality is on average more important because the products are intended for either consumption or to improve consumer health. Notably, although Harbingers are generally less likely to purchase healthcare products, when they do purchase them it is a particularly strong signal that the product will fail.

We next compare the outcomes for national brands and private label products. Our sample of 2,953 new products includes 18% private label products (the retailer's store brand). The results from re-estimating the models separately for private label and national brand new products in the prediction set are also reported in the Appendix (Table A16). Again, we observe a negative effect for Group 4 Sales, which replicates the harbinger effect. We also find that the improvement in predictive performance (measured by the area under the ROC curve) is particularly strong for private label items.

In the Appendix (Tables A17 and A18) we also compare the results across products with different price levels and discount intensities. We used median splits to divide the sample of 2,953 new products into low- and high-priced items based on average prices, and into less- and more-frequently discounted items based on percentage of sales on promotion.

We also investigated whether the findings varied according to how "risky" the new product was. To do so we identified whether the product introduced a brand to the product category, or whether it was introduced in a product category with high average failure rates. However, we did not find large differences when distinguishing between the new products in these ways.

The comparisons across product categories demonstrate the robustness of the effect. Re-estimating the model on these separate samples of products serves as a replication check, confirming that the effect is not limited to a small subset of items or categories.

Summary

We have presented evidence that early adoption of a new product by some groups of customers is associated with a higher probability that the new product will fail. The findings survive a range of robustness checks.

In the next section we ask: Who are the Harbingers? In particular, we compare their purchasing behavior with other customers. This leads us to investigate whether we can identify Harbingers through their purchases of existing products.

Who Are the Harbingers?

To help characterize who these customers are, we divide the customers into “Harbingers” and “Other” customers based on their classification set purchases. Harbingers include customers in Groups 3 and 4, while the Other customers are in Groups 1 and 2. In Table 3 we compare the purchasing patterns of the two groups of customers using the transactions in the period used to identify the classification set (November 2003 to July 2004). We include purchases of all products (new and existing) and in the Appendix (Table A21) we repeat the analysis when focusing solely on new products. Definitions and summary statistics of these purchasing measures are also provided in the Appendix (Tables A19 and A20).

Table 3

The findings in Table 3 reveal that on average Harbingers purchase more items but visit a similar number of stores. They tend to buy slightly more items per visit, but make slightly fewer visits. Although statistically significant, the differences in these measures are relatively small. There are larger differences in the prices of the items that they purchase and the categories that they purchase from. Harbingers tend to choose less expensive items and are more likely to purchase items on sale and items with deeper discounts. They purchase a higher proportion of beauty items, but a lower proportion of healthcare items.

Harbingers tend to purchase new products more quickly after the items are introduced (see Table A21 in the Appendix). On average they purchase new products 26.8 weeks after they are introduced, compared to 27.9 weeks for other customers. The tendency of Harbingers to purchase new products slightly earlier may mean that we observe a slightly higher proportion of Harbingers purchasing during the initial evaluation periods. However, this cannot explain the findings that we reported in the previous section as this affects all new products (not just the new products that fail). Other comparisons of the purchases of new products (see Table A21 in the Appendix) reveal an almost identical pattern to the purchase of all products (Table 3).

Preference Minorities

Although our data is not well-suited to conclusively explaining why purchases by Harbingers signal that a new product is likely to fail, in the Introduction we speculated that Harbingers may have product preferences that are different from the general population. If this is the case, when a Harbinger adopts a new product, it may signal that the product is not a good match for the preferences of other customers. This explanation is related to previous work on "preference minorities". Recall that Choi and Bell (2011) investigate variation in the adoption of online shopping across different geographies. They show that customers whose preferences are not representative of other customers in the area are more likely to purchase online, presumably because local offline retailers have tailored their assortments to other customers (see also Waldfogel 2003).

We can investigate this explanation by asking whether customers with high *FlopAffinities* are also more likely to purchase *existing* products that other customers do not buy. Using the aggregate store transaction data, we calculate *Total Unit Sales* for each item sold in the 111 stores in calendar year 2008 (focusing on existing products by excluding the new products). We then order the items according to *Total Unit Sales* and define an item as a "niche" or "very niche" product if it is among the items that contribute the fewest units sold. Niche items collectively contribute 1% of total unit sales, while very niche items collectively contribute just 0.1% of total unit sales. We then average across each customer's item purchases to calculate the following three measures:

Unit Sales	The average of *Total Unit Sales*.
Niche Items	The proportion of items that are niche items.
Very Niche Items	The proportion of items that are very niche items.

When averaging across each customer's purchases, we weight the items using the number of units of that item purchased by that customer.[11] The findings are reported in Table 4, where for ease of comparison (and to protect the confidentiality of the company's data), we scale the measures to 100 for customers in Group 1.

The findings reveal a clear pattern: customers in the highest *FlopAffinity* groups are much more likely to purchase items that few other customers purchase. Customers in Group 4 purchase items that sell over 9% fewer total units than customers in Group 1. They also purchase 9% more *niche* items and 12% more *very niche* items. For all three measures the differences between the Harbingers (Groups 3 and 4) and the other groups are statistically significant ($p < 0.01$).

Table 4 about here

When interpreting these findings it is important to recall that this comparison focuses exclusively on existing items (we exclude the new products in the classification and prediction sets). If the analysis were conducted on new products it would seem unsurprising that customers who buy niche products are customers who are more likely to buy products that fail. What the findings in Table 4 reveal is that Harbingers not only purchase new products that do not succeed; they are also more likely to purchase existing products that have relatively low sales.

This result is consistent with an explanation that Harbingers have preferences that are systematically different from other customers. If Harbingers adopt a new product, it may

[11] The findings are robust to weighting each product equally.

signal that other customers will not be attracted to the product. This is essentially the opposite of the argument that von Hippel (1986) proposes for why firms can benefit by distinguishing 'lead users' from other customers. While lead users provide a positive signal of product success, Harbingers provide the opposite signal.

The findings in Table 4 also suggest another mechanism that firms can use to identify Harbingers. Recall that in our analysis we identified Harbingers using purchases of new products in the classification set. The results in Table 4 suggest that we may also be able to identify Harbingers using purchases of existing products. In particular, we can classify customers according to whether they purchased niche or very niche (existing) products. In Table 5 we report the findings when using these customer groupings to predict the success of the new products in the prediction set.

Table 5 about here

The findings confirm that customers who purchase niche or very niche (existing) products are also Harbingers (of product failure). Purchases of new products by these preference minorities provide an additional signal that the new product will fail. Comparing the AUC measures in Columns 2 and 3 with the base model (Column 1) indicates that the *FlopAffinity* and a tendency to purchase niche products provide similar predictive information. Moreover, the two signals provide independent information. The predictive power of the model when including both signals (Column 4) is greater than when using just one of these approaches (Columns 2 and 3). This indicates that the measures do not perfectly coincide; not all customers who have a high *FlopAffinity* purchase niche products, and vice versa. We conclude that both types of customers can be considered Harbingers, whose adoption of new products signals product failure.

Summary

We have shown that Harbingers are more likely to purchase niche items that other customers do not purchase, indicating that they have preferences that are less representative of mainstream tastes. This insight suggests that purchases of existing items may also be used to identify Harbingers. Further investigation confirms that adoption by customers who tend to purchase niche (existing) items also provides a signal that a new product will fail.

Conclusions

Using a comprehensive dataset from a large retail chain, we have shown that the early adoption of a new product by different groups of customers provides different signals about the likelihood that a product will succeed. In particular, there exist "Harbingers" of failure*:* customers whose decision to adopt a new product is a signal that the product will fail. The signal is even stronger if these customers not only adopt the products, but they also come back and purchase again. We present evidence that Harbingers have preferences that are not representative of other customers in the market, and that a pattern of adoption of niche products represents an alternative way of identifying them from data.

The findings have an important managerial implication. They suggest that not all early adopters of new products are the same. For some customers, adoption of a new product is an indication that the product is more likely to succeed. However, for Harbingers, adoption is an indication that the product will fail. When firms use early adoption to make product line decisions or as input to the product improvement process, it is important to distinguish between these different types of customers.

There are two important limitations to this research. First, we have demonstrated the Harbinger effect using data from a single retailer that sells consumer packaged goods. Replicating the findings using data from different firms and in other categories will be important to confirm the generalizability of the findings. Second, our investigation has focused on showing that Harbingers have preferences that are not representative of other customers. However, we cannot determine whether these unusual preferences are endowed, learned, or in general, where they come from. Moreover, although we show that our two approaches to identifying Harbingers (past purchases of new product failures and purchases of existing products that are niche or very niche) both have independent predictive value, it is unclear why this is the case. Additional research is required to determine whether they provide separate information about the same construct (e.g. non-representative preferences), or whether they provide information about two distinct constructs.

Future research could also address the challenge of recognizing which customers are Harbingers. Our retail setting, where we can track purchases of different products by a panel of individual customers, provides one mechanism for doing this. However, in other settings without a sequence of individual transactions, other mechanisms may be required to identify these customers. The evidence that Harbingers are more likely to purchase existing products that few other customers purchase may provide useful clues even without access to detailed purchase histories. Finally, while our results provide convergent evidence of Harbingers of Failure, we also have some evidence that there may be Harbingers of Success. Future research is needed to more accurately identify both types of Harbingers.

References

Anderson, Evan E. (1979), "An Analysis of Retail Display Space: Theory and Methods," *Journal of Business*, 52(1), 103-118.

Barbier, Joel, Mary de Wysocki, Steve Du Mont, Sharon Finke, Shaun Kirby, James Macaulay, Mike McLaughlin, Waseem Sheikh, Jon Stine, and Ronald van Zanten (2005), "Improving Innovation Productivity in the Consumer Packaged Goods Industry," working paper, Cisco Internet Business Solutions Group.

Bass, Frank M. (1969), "A New Product Growth Model for Product Diffusion," *Management Science*, 15(5), 215-227.

Biyalagorsky, Eyal, William Boulding, and Richard Staelin (2006), "Stuck in the Past: Why Managers Persist With New Product Failures," *Journal of Marketing*, 70(2), 108–121.

Boulding, William, Ruskin Morgan, and Richard Staelin (1997), "Pulling The Plug To Stop The New Product Drain," *Journal of Marketing Research*, 34(1), 164–176.

Brockner, Joel (1992), "The Escalation of Commitment to A Failing Course of Action," *Academy of Management Review*, 17(1) 39–61.

Brockner, Joel, Jeffrey Z. Rubin (1985), *Entrapment in Escalating Conflicts: A Social Psychological Analysis*, Springer-Verlag, New York.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Mohammad S. Rahman (2009), "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," *Management Science*, 55(11) 1755–1765.

Brynjolfsson, Erik, Yu (Jeffrey) Hu and Duncan Simester (2011), "Goodbye Pareto Principle, Hello Long Tail: The Effect of Internet Commerce on the Concentration of Product Sales," *Management Science*, 57(8), 1373-1387.

Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49(11), 1580–1596.

Choi, Jeonghye and David R. Bell (2011), "Preference Minorities and the Internet," *Journal of Marketing Research*, XLVIII, August, 670-682.

Eskin, Gerald J. (1973), "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," *Journal of Marketing Research*, 10(2), 115-129.

Eskin, Gerald J. and J. Malec (1976), "A Model for Estimating Sales Potential Prior to the Test Market," *Proceedings of the American Marketing Association 1976 Fall Educators' Conference*, American Marketing Association, Chicago IL, 230-33.

Fader, Peter S., Bruce G. S. Hardie (2005), "The Value of Simple Models in New Product Forecasting and Customer-Base Analysis," *Applied Stochastic Models in Business & Industry*, 21(4/5), 461-473.

Henard, David H. and David M. Szymanski (2001), "Why Some New Products are More Successful than Others," *Journal of Marketing Research* 38(3), 362-375.

Huang, Yan, Param V. Singh, and Kannan Srinivasan (2014), "Crowdsourcing New Product Ideas under Consumer Learning," *Management Science*, 60(9), 2138-2159..

Lilien, Gary L., Pamela D. Morrison, Kathleen Searls, Mary Sonnack, and Eric von Hippel (2002), "Performance Assessment of The Lead User Idea-Generation Process for New Product Development," *Management Science*, 48(8), 1042-1059.

Liutec, Carmen, Rex Du and Ed Blair (2012), "Investigating New Product Purchase Behavior: A Multi-Product Individual-Level Model for New Product Sales," working paper, University of Houston.

Mahajan, Vijay, Eitan Muller and Frank M. Bass (1990), "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54(1), 1-26.

Moore, Geoffrey A. (1991), *Crossing the Chasm*, HarperBusiness, New York.

Morvinski, Coby, On Amir and Eitan Muller (2014), " 'Ten Million Readers Can't Be Wrong!', or Can They? On the Role of Information about Adoption Stock in New Product Trial," working paper.

Parfitt, J. H. and B. J. K. Collins (1968), "Use of Consumer Panels for Brand-Share Prediction," *Journal of Marketing Research*, 5(2), 131-145.

Pringle, Lewis G., Dale Wilson and Edward I. Brody (1982), "NEWS: A Decision Oriented Model for New Product Analysis and Forecasting," *Marketing Science*, 1(1), 1-30.

Simester, Duncan and Juanjuan Zhang (2010), "Why Are Bad Products So Hard to Kill?" *Management Science*, 56(7), 1161-1179.

Silk, Alvin J. and Glen L. Urban (1978) "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *Journal of Marketing Research*, 15(2), 171-191.

Urban, Glen L. and John R. Hauser (1993), *Design and Marketing of New Products*, Prentice Hall, Englewood Cliffs NJ.

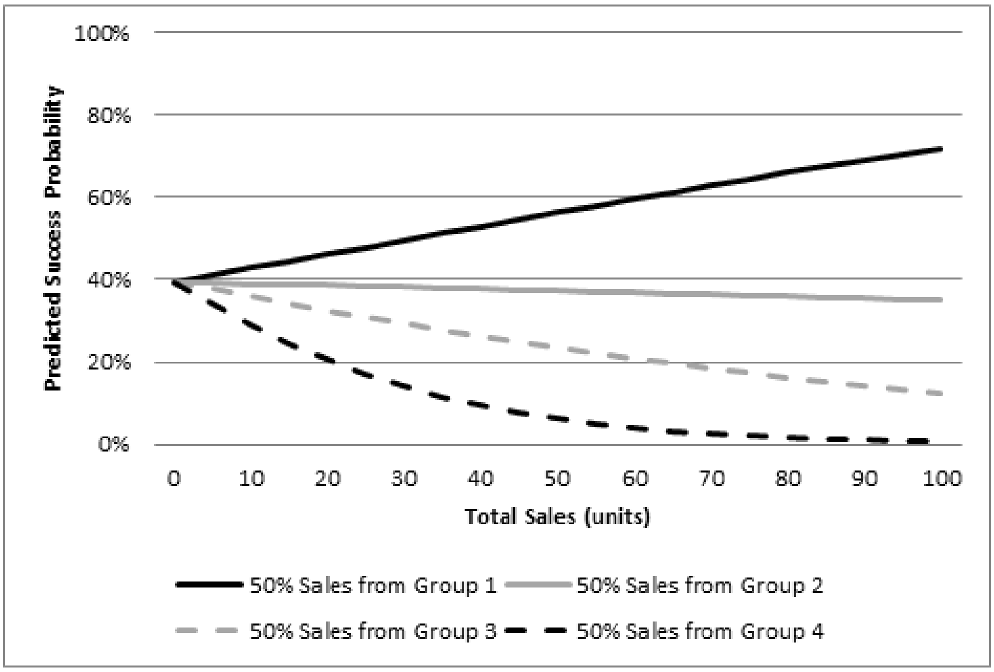
Urban, Glen L. and Eric von Hippel (1988), "Lead User Analyses for the Development of New Industrial Products," *Management Science*, 34(5), 569-582.

Van den Bulte, Christophe and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26(3), 400-421.

von Hippel, Eric (1986), "Lead Users: A Source of Novel Product Concepts," *Management Science*, 32(7), 791-805.

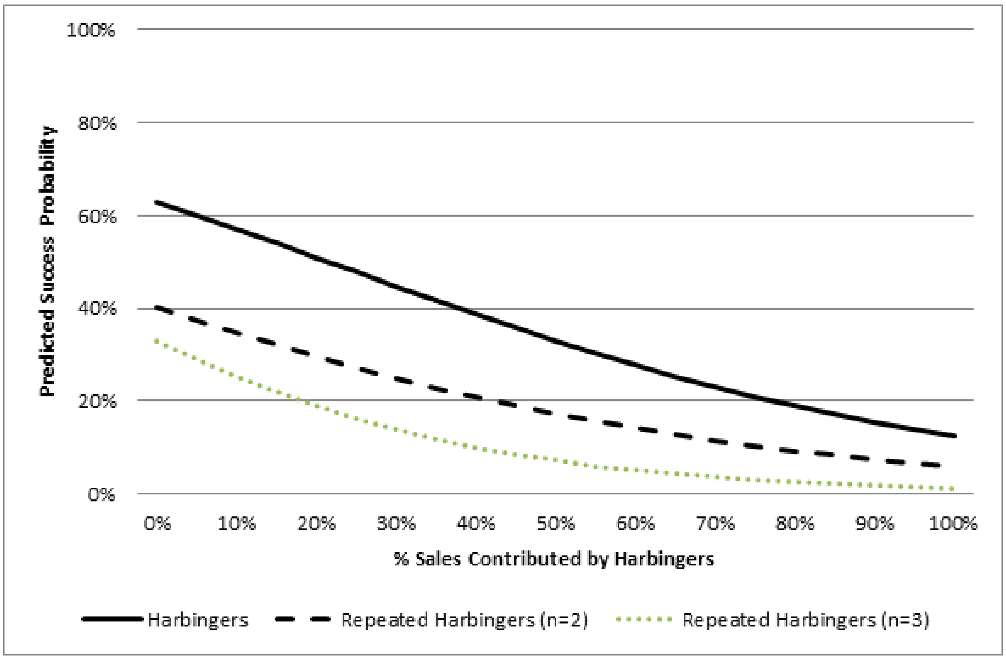
Waldfogel, Joel (2003), "Preference Externalities: An Empirical Study of Who Benefits Whom in Differentiated Product Markets," *RAND Journal of Economics*, 34(3), 557-568.

Figure 1. Predicted Success Probability as Total Sales Increase



The figure reports the predicted probability that a new product will succeed using the calibrated parameters from Model 2 of Table 1. Each curve represents how predicted success changes as sales increase, assuming 50% of the sales come from one of the four groups and the remaining 50% of sales is distributed equally across the other three groups and all other unclassified customers.

Figure 2. Predicted Success Probability as Sales from Harbingers Increase



The figure reports the predicted probability that a new product will succeed using the calibrated parameters from Model 2 of Table 2. The sales volume is fixed at the empirical average of the sample. Each curve represents how the probability of success varies as the percentage of sales from Harbingers increases. The solid curve is generated from the model that defines Harbingers as customers for which *FlopAffinity* is between 0.5 and 1. The two dashed curves are generated from a model that defines Harbingers as customers for which *Repeated* $FlopAffinity_i$ (n=2 and 3) is between 0.5 and 1.

Table 1. Marginal Effects From Logistic Models

	Model 1	Model 2	Model 3	Model 4
Total Sales	0.0011**		0.0025**	
	(0.0004)		(0.0006)	
Group 1 Sales		0.0113*		0.0056
		(0.0049)		(0.0041)
Group 2 Sales		0.0016		0.0004
		(0.0055)		(0.0050)
Group 3 Sales		-0.0067		-0.0018
		(0.0036)		(0.0032)
Group 4 Sales		-0.0258**		-0.0165**
		(0.0052)		(0.0048)
Sales from Other Customers		0.0114**		0.0098**
		(0.0023)		(0.0021)
No Sales in the first 15 weeks			0.1156	0.1037
			(0.0769)	(0.0761)
(log) Price Paid			0.0500**	0.0432*
			(0.0181)	(0.0180)
Profit Margin			0.0300	0.0259
			(0.1226)	(0.1191)
Discount Received			-0.0389	-0.0105
			(0.1604)	(0.1552)
Discount Frequency			-0.1187	-0.1173
			(0.0951)	(0.0953)
Herfindahl Index			0.1996	0.2134*
			(0.1113)	(0.1034)
Category Sales			-0.1025**	-0.0979**
			(0.0340)	(0.0338)
Vendor Sales			-0.0304	-0.0317
			(0.0334)	(0.0338)
Private Label			0.2499**	0.2362**
			(0.0464)	(0.0469)
Nbr. Customers with 1 repeats			-0.0134	-0.0063
			(0.0081)	(0.0092)
Nbr. Customers with 2 repeats			-0.0390	-0.0423
			(0.0199)	(0.0240)
Nbr. Customers with 3 or more repeats			0.0038	0.0181
			(0.0288)	(0.0413)
Log Likelihood	-1,998	-1,952	-1,823	-1,800
Likelihood Ratio Test, Chi^2 (df=4)		90.24**		46.66**
Area under ROC curve	0.6035	0.6160	0.7104	0.7242

The table reports average marginal effects from models where the dependent variable is a binary variable indicating whether the new product succeeded (1 if succeeded, 0 if failed). Robust standard errors (clustered at the category level) are reported in parentheses. The unit of analysis is a new product and the sample size is 2,953 new products. Significantly different from zero (or significant difference between Models 1 and 2): * $p<0.05$, ** $p<0.01$.

Table 2. Grouping Customers by Repeat Purchases

	Model 1	Model 2		
		At Least 1 Purchase	At Least 2 Purchases	At Least 3 Purchases
Total Sales	0.0011** (0.0004)			
Groups 1 and 2		0.0064* (0.0032)	-0.0058 (0.0042)	-0.0096 (0.0087)
Groups 3 and 4		-0.0144** (0.0027)	-0.0218** (0.0051)	-0.0375** (0.0090)
Sales to Other Customers		0.0118** (0.0022)	0.0066** (0.0012)	0.0045** (0.0008)
Log Likelihood	-1,998	-1,959	-1,963	-1,962
Likelihood Ratio Test, Chi^2 (df=2)		78.69**	69.57**	73.06**
Area under ROC curve	0.6035	0.6128	0.6050	0.6159

The table reports average marginal effects from models where the dependent variable is a binary variable indicating whether the new product succeeded (1 if succeeded, 0 if failed). Robust standard errors (clustered at the category level) are reported in parentheses. The unit of analysis is a new product. The sample size is 2,953. Significantly different from zero (or significant difference between Models 1 and 2): * $p<0.05$, ** $p<0.01$.

Table 3. Harbingers vs. Others Purchasing of *New and Existing* Products

	Harbingers	Others	Difference
Total Purchases	75.679	73.588	2.091* (1.023)
Purchases per Visit	4.026	3.750	0.277** (0.022)
Purchases per Store	34.249	34.774	-0.525 (0.484)
Shopping Visits	20.204	20.912	-0.768** (0.248)
Different Stores Visited	2.841	2.811	0.030 (0.030)
Regular Price of Items	$4.378	$4.792	-$0.414** (0.022)
Price Paid	$3.716	$4.181	-$0.464** (0.021)
Discount Received	14.449%	12.944%	1.505%** (0.123%)
% Discounted Items	35.239%	33.327%	1.912%** (0.232%)
% Beauty Items	18.591%	16.054%	2.537%** (0.170%)
% Edible Items	30.367%	29.101%	1.266%** (0.205%)
% General Merchandise Items	22.350%	21.449%	0.900%** (0.157%)
% Health Items	21.680%	25.991%	-4.311%** (0.183%)

The table reports the purchasing behaviors for both Harbingers and Other customers. All measures are calculated using purchases of both new and existing products in the classification period of the transaction data (November 2003 to July 2004). Standard errors of the mean difference are reported in parentheses. The sample size is 29,463. Harbingers are customers from Groups 3 & 4 (n = 16,620), while Others are customers from Groups 1 & 2 (n = 12,843). Significantly different from zero: * p<0.05, ** p<0.01.

Table 4. Purchases of Niche Products

FlopAffinity	Unit Sales	Niche Items	Very Niche Items
Group 1	100.00 (1.66)	100.00 (0.36)	100.00 (0.45)
Group 2	90.26 (1.56)	104.78 (0.39)	106.24 (0.48)
Group 3	91.98 (1.22)	105.66 (0.30)	107.25 (0.38)
Group 4	90.49 (1.73)	109.18 (0.40)	111.76 (0.51)

The table reports the average of each measure by customer group (standard errors are in parentheses). The measures are initially calculated for each customer, and then averaged across customers within each group. The groups are defined using the *FlopAffinity* in customers' classification sets. For ease of interpretation (and confidentiality reasons) we index the measures to 100 in Group 1. The sample size is 29,412, less than the full sample size 29,436 because 51 customers did not purchase any existing products in the classification period.

Table 5. Incorporating Preference Minorities

	Model 1	Model 2	Niche Purchasers	Both Measures
Total Sales	0.0011**		0.0050**	
	(0.0004)		(0.0012)	
Group 1 Sales		0.0113*		0.0105*
		(0.0049)		(0.0049)
Group 2 Sales		0.0016		0.0052
		(0.0055)		(0.0057)
Group 3 Sales		-0.0067		-0.0017
		(0.0036)		(0.0041)
Group 4 Sales		-0.0258**		-0.0217**
		(0.0052)		(0.0053)
Sales to Other Customers		0.0114**		0.0116**
		(0.0023%)		(0.0022)
Sales to Customers Buying Niche Items			0.0066	0.0114
			(0.0068)	(0.0068)
Sales to Customers Buying Very Niche Items			-0.0284**	-0.0226**
			(0.0070)	(0.0070)
Log Likelihood	-1,998	-1,952	-1,969	-1,943
Likelihood Ratio Test, Chi^2 (df = 2 or 4)		90.24**	57.49**	20.01**
Area under ROC curve	0.604	0.616	0.617	0.629

The table reports average marginal effects from models where the dependent variable is a binary variable indicating whether the new product succeeded (1 if succeeded, 0 if failed). Robust standard errors (clustered at the category level) are reported in parentheses. The unit of analysis is a new product. The sample size is 2,953. The Chi^2 test compares Model 1 to Models 2 (df = 4) and 3 (df = 2), and Model 2 to Model 4 (df = 2). Significantly different from zero (or significant difference between Models 1 and 2): * *p*<0.05, ** *p*<0.01.