The Effects of Personalized Recommendations with Popularity Information on Sales - A Field Study in Grocery Retailing

Marc Linzmajer  
*University of St. Gallen*, marc.linzmajer@unisg.ch

Sandro Schopfer  
*ETH Zürich*, sandro.schopfer@ethz.ch

Thorben Keller  
*University of St. Gallen*, thorben.keller@unisg.ch

Liane Nagengast  
*University of St. Gallen*, liane.nagengast@unisg.ch

Elgar Fleisch  
*University of St. Gallen*, elgar.fleisch@unisg.ch

*See next page for additional authors*

Follow this and additional works at: [http://aisel.aisnet.org/ecis2015_rip](http://aisel.aisnet.org/ecis2015_rip)
Authors
Marc Linzmajer, Sandro Schopfer, Thorben Keller, Liane Nagengast, Elgar Fleisch, and Thomas Rudolph
THE EFFECTS OF PERSONALIZED RECOMMENDATIONS WITH POPULARITY INFORMATION ON SALES – A FIELD STUDY IN GROCERY RETAILING

Research in Progress

Linzmaner, Marc, University of St. Gallen, Switzerland, marc.linzmaner@unisg.ch
Schopfer, Sandro, ETH Zurich, Switzerland, sandro.schopfer@ethz.ch
Keller, Thorben, University of St. Gallen, Switzerland, thorben.keller@unisg.ch
Nagengast, Liane, University of St. Gallen, Switzerland, liane.nagengast@unisg.ch
Fleisch, Elgar, University of St. Gallen, Switzerland, elgar.fleisch@unisg.ch
Rudolph, Thomas, University of St. Gallen, Switzerland, thomas.rudolph@unisg.ch

Abstract

In consumer and information systems research, it remains unclear how consumers consider smartphone app recommendations in the course of their decision making process that leads to product choices in the physical store. Moreover, it is unclear which type of information smartphone apps should transport to consumers and if there are any customer segmentation criteria for smartphone app design. With respect to the theoretical and managerial importance of recommendation services in the form of smartphone apps we want to shed some light on this topic. Combining literature from the fields of IS and marketing research, we hypothesize that personalized recommendations via smartphone apps can help to boost sales in physical grocery stores. Furthermore, we hypothesize that additional popularity information (in the form of “stars”) does not amplify the positive effect of personalized recommendations. In addition, we assume that the effects of recommendation usage differ for men and women. We conducted a field study with a European grocery retailer to test our hypotheses. Finally, we discuss first implications as well as central limitations of our research and present the next research steps.

Keywords: Personalized Recommendations, Popularity Information, Mobile Apps, Shopper Marketing, Gender Differences.

1. Introduction

The evolution of information technologies (IT) has changed the way firms are adapting to consumers’ needs. In the past two decades, research results have led to important implications for information systems (IS) and marketing managers with regard to the development of new services and advertising content employed to increase consumers’ intentions to purchase products in online stores. Researchers have studied the factors influencing the acceptance of these new technologies (e.g., Davis et al., 1989; Venkatesh and Davis, 2000; Venkatesh et al., 2003), the characteristics of online consumers (e.g., Holzwarth et al., 2006; Gefen and Straub, 2003; Koufaris, 2002; Pavlou, 2003), and the determinants of online purchasing behavior (e.g., Gefen et al., 2003; Rodríguez-Ardura et al., 2008). Not only for products sold online, but also for physical products sold in-store, mass customization and fast response to dynamic needs have become crucial to remaining competitive (see Shankar et al., 2011). In this context, mobile smartphone application (in the following: smartphone app) usage has evolved as an important research topic in the field of management IS. As smartphone apps provide new opportunities
for recommendation systems, some retailers have started to implement recommendation agents to target their customers with personalized offers. However as from a theoretical as well as from a practical perspective, these mobile recommendation systems are still in their infancy, IT- and marketing managers need to gain a deeper understanding about how to design recommendation agents, which customers benefit most, and which specific needs should be addressed.

Understanding the shopping behavior of smartphone app users is essential to decide whether or not the implementation of smartphone apps in general and recommendation agents in particular will be successful. Therefore, the question of how smartphone app usage affects the purchase behavior in retail stores is at the heart of marketing and IS research. To date it is unclear how smartphone app recommendations affect consumers’ decision making processes that lead to product choices in the physical store. First, it is questionable, which type of information smartphone apps should transport to consumers to foster in-store sales. Second, it is unclear if all consumer segments react identically to product recommendations or if there are any segmentation criteria for the design of smartphone app recommendations. For example in a sole online context, gender has been identified as important segmentation variable because men and women tend to perceive recommendation messages differently (Massar and Buunk, 2013; Wolin, 2003). Against this background, the objective of our study is to examine the impact of different personalized recommendations (i.e., with and without popularity information) through smartphone apps on purchase decisions of female and male customers.

With respect to the theoretical and managerial importance of recommendation services provided through smartphone apps, we aim to shed some light on this topic by answering the following research questions:

1. Do consumers who receive personalized recommendations via their smartphone spend more money on their in-store grocery purchase trip compared to consumers who do not receive such recommendations?
2. Does the integration of popularity information in personalized recommendation via smartphones affect this relationship?
3. Do men and women react differently to personalized recommendations via smartphones?

In what follows, we first derive our research hypotheses based on prior literature, and then describe the methodology to analyze the hypothesized effects. After presenting the results of a field study in European grocery retailing, we discuss first implications and conclude with important limitations and further research that is planned within this project.

2. Literature and Research Hypotheses

Existing research has examined the influence of interpersonal communication in face-to-face settings, under personal influence or word-of-mouth (e.g., Duhan et al., 1997; Gilly et al., 1998; Gershoff et al., 2001; Rosen and Olshavsky, 1987). With the advent of new IT, a recommender today might appear in the form of a recommendation system on a mobile device, for instance through a smartphone app. A recommendation or recommender system describes an information system that provides content or product information online to meet the needs of a particular customer (Liang et al., 2007, p. 47). Due to the expansion of the internet, a new research area has emerged in the fields of IT as well as consumer behavior, namely that of impersonal sources that provide personalized information (e.g., Alba et al., 1997; Anasari et al., 2000; Komiak and Benbasat, 2004). Based on the literature on consumers’ pre-purchase information search, advantages of recommendation systems mainly arise from the principle of least effort and information overload: The principle of least effort states that each individual will adopt a course of action that will involve the least average work from the person. This principle that is supported by evidence from different studies of language usage (Zipf 1949), predicts that information seekers will minimize the effort required to obtain information, even if it means accepting a lower quality or quantity of information (Allen 1977). From several studies, it is evident that accurate content recommendation, which reduces the effort needed by a user to search for relevant product information, can increase user satisfaction and thereby facilitate purchases decisions (Senecal and Santel, 2004).
Combining recommendations with personalization, Liang et al. (2007) give an overview of different theoretical accounts related to personalized content services: They show that in IS and retailing research, personalized recommendations are mainly seen as a customized information source to reduce the customer’s effort in making the right product choices. This mechanism is especially important with regard to food products in grocery stores, because this category is characterized by a broad range of products with similar functional properties, leading to information overload and consumer confusion (Huffman and Kahn, 1998; Balabanis and Craven, 1997; Schweizer et al. 2006; Walsh et al., 2007). This means consumers are given more information that they can handle within a given time frame (Liang et al., 2007). Through personalized recommendation agents (e.g. via smartphone apps), IT might be useful in alleviating information overload (e.g., through information customization during grocery shopping).

Based on this general understanding of personalized recommendations, we screen the literature more focused with regard to its effects on consumer purchase decisions and derive hypotheses in the following.

2.1 Personalized recommendations and purchase decisions

Past empirical research has shown that personal and impersonal information sources influence consumers’ decision-making (e.g., Gilly et al., 1998). More specifically and applied to interactive contexts, the adoption of recommendation agents in online contexts is determined by perceived personalization and moderated by specific forms of trust (Komiak and Benbasat, 2006). In addition, Häubl and Trifts (2000) have shown that interactive tools such as recommendation agents (like smartphone apps in our study) have strong positive effects on both the quality and the efficiency of purchase decisions in an online environment. As electronic marketplaces present consumers with diverse conditions fostering considerable uncertainty, consumers try to make purchase decisions that reduce this uncertainty and recommendation tools might be helpful in this sense (Häubl and Trifts, 2000). Correspondingly, several studies revealed that recommendation agents help to reduce consumers’ information overload (Todd and Benbasat, 1999), improve decision quality (Pereira, 2001), and finally, influence consumer behavior and purchase intentions (Bo and Benbasat, 2007; Kamis et al., 2008; Kowatsch and Maass, 2010). The economic benefit of recommendations has been shown in an experimental design by Senecal and Nantel (2004): consumers who consult online product recommendations selected recommended products twice as often as consumers who did not consult recommendations. Finally and based on the literature on consumers’ pre-purchase information search, we believe that personalized recommendations will have a greater influence on consumers’ purchase decisions than non-personalized ones (Brown and Reingen, 1987; Kim et al., 2002). We assume that these results from online buying-contexts can be transferred to offline contexts that are combined with recommendations from a smartphone app (for an overview of shopping behaviors in online and offline channels for grocery products, see Chu et al., 2010). As products in grocery retailing are often characterized by similar functional properties, we assume that this product characteristic underlines the importance of recommendation agents to reduce information overload and consumer confusion (Huffman and Kahn, 1998; Balabanis and Craven, 1997; Schweizer et al. 2006; Walsh et al., 2007; Van der Heijden, 2006). Therefore, we hypothesize:

H1. Consumers, who receive personalized recommendations via their smartphone, will spend more money on their in-store purchase trip compared to consumers, who do not receive personalized recommendations.

2.2 Personalized recommendations, popularity information and purchase decisions

From a managerial viewpoint, it is intuitively reasonable to combine the positive effects of personalized recommendations with popularity information. Literature on observational learning shows that decision makers tend to follow peer choices as they infer product quality from what their peers have chosen (Banjeree, 1992). Empirical studies in this domain have emphasized evidence of quality inference, either
in the lab (Celen and Kariv, 2004) or in the field (Zhang, 2010; Chen et al., 2011). These conclude in general, that popularity information benefits high volume items. In a more fine-grained setting, Tucker and Zhang (2011) show that this conclusion does not hold over all conditions and demonstrate cases where popularity information does not pay off. In the context of personalized recommendations, we hypothesize that popularity information is not always perceived favorable and that it might even cause feelings of cognitive dissonance. Festinger (1962) introduced the theory of cognitive dissonance, which “[…] centers around the idea that if a person knows various things that are not psychologically consistent with one another, he will, in a variety of ways, try to make them more consistent” (Festinger 1962, p. 93). When a consumer receives a personal recommendation from a retailers’ recommendation system, he or she won’t expect any popularity information as it contradicts the factor “personal” and, hence, the individuality of the recommendation. This leads to our assumption that a dissonance reduction appears in the form of spending less money on recommended products with popularity information. In summary, popularity information might diminish the credibility of personal recommendations through a cognitive-dissonance effect. More formally, we hypothesize:

H$_3$. Consumers, who receive personalized recommendations combined with popularity information via their smartphone, might spend less money on their in-store purchase trip compared to consumers, who receive personalized recommendations without popularity information.

2.3 Gender differences in the effect of personalized recommendations, popularity information and purchase decisions

Gender is one of the key attributes and predictors of online purchase behaviors. In the IT-discipline, an extensive number of empirical studies document gender differences in general areas such as the use of computers and the Internet, but also in more specific areas such as online trust and related behaviors (for an overview see e.g. Riedl et al., 2010; Okazaki, 2007). Shopping in general plays a more emotionally encompassing role for women than for men (Campbell, 2000): women have highly positive attitudes toward shopping, associating it with a leisure, whereas men tend to have more negative attitudes toward buying, viewing it as work that should be accomplished with a minimum input of time and effort. Women, therefore, tend to focus on the enjoyable process of buying, whereas men primarily focus on the outcome of obtaining the goods (Dittmar et al., 2004). Moreover, men are more functional in their buying attitudes than women, who, in turn, are more inclined to emphasize emotional concerns (Dittmar et al., 1996). With respect to online shopping, male buyers are more convenience-oriented and less motivated by social interaction than female buyers (Swaminathan et al., 1999). Another study (Awad and Ragowsky, 2008) investigated the effect of gender on the relationship between online word-of-mouth quality and online trust. The results of this study reveal that the effect of online trust on intention to buy online is stronger for women than for men. Given the research on gender differences in the IT realm, and gender differences in online trust and related behaviors, evidence supporting substantial behavioral differences between women and men is available. Combining this research, more functional-oriented men are likely to see personalized recommendations on their smartphone app as decision aid in a complex grocery retail environment that reduces uncertainty. On the other hand, it is harder for women to trust impersonal decision aids like personalized mobile recommendations as their shopping decisions are more emotionally driven. Considering this differential psychological mechanism, we have reason to state the following prediction:

H$_3$. When receiving in-store personalized recommendations (with or without popularity information) via their smartphone, men might spend more money on recommended products compared to women.

In the following, we describe our methodology followed by a discussion of the first implications of our research and an outlook on further research that is planned.
3. Methods

3.1 Field design and stimulus material

In order to answer our research questions we used a field study design (for a discussion of field experiments, see Harrison and List, 2004) and have developed a mobile application that is connected to the retailer’s point-of-sale system (POS) and thus can be used to push personalized product recommendations. The recommendation engine analyses past purchase behavior and determines suitable product alternatives using a simple basket-analysis approach. Every product of the assortment is classified into 1 out of 913 different categories $C$, where the list of alternatives is defined to be the set of products also contained in $C$. App users are able to access a feature called personal recommendations where the recommendations are presented as depicted in figure 1. Based on the sales rank of a product in a given category, popularity information was added in the form of golden stars (top 10%) and silver stars (top 30%). No stars were shown for the remaining products. Apart from the popularity information the name of the product, the producer and, where available, a picture of the product is shown.

![Figure 1 Personal product recommendations with popularity information](image)

The app was launched in a single store where some advertisement was placed to promote it. There was no pre-selection process implemented and every customers was free to download and use the system. Half of the app users did not see any recommendations at all (group “NoRecomm”) and the rest was divided into two groups “StarSeen” and “NoStarsSeen”. Users in the group “NoRecomm” serve as the control group for our experiments. As the name suggests only users in the group “StarSeen” were able to see the popularity information while they were hidden from users in the group “NoStarsSeen”.

Using the implemented tracking engine we were able to exactly analyze which customer saw which recommendations and when exactly this happened. The connection to the POS then allowed us to determine if a customer bought a specific product after it was recommended to him or her.

3.2 Data collection and analysis

This section describes how the transaction data is processed using a Bayesian approach. The processing is done in two major steps. First, the posterior distributions of the mean amount spent for all experimental groups are computed. Using the posteriors, the probability that participants of an experimental group have spent more on average compared to participants of another experimental group can be calculated as a second step. The methodology is attractive because of its fast processing power and the ability to graphically compare posteriors of the mean amount spent over all experimental groups. Over the course of this field study a number of digital receipts have been collected using the smart phone application distributed over a total of 69 participants. Over all experimental groups a total of 408 digital receipts have been collected. The value (i.e. the amount that some participant spent in EUR) of the i-th digital receipt is denoted as $x_i$. We found that the distribution of the amount paid matches well the exponential distribution $f(x) = \exp(-x/\mu)/\mu$ for all receipts in an experimental group. Instead of
trying to estimate the average amount paid $\mu$ directly we make use of Bayes theorem to compute the distribution of the parameter $\mu$ conditioned on the data collected. Using a suitable prior $g(\mu)$ for the parameter $\mu$ the posterior (i.e., the distribution of the parameter $\mu$ given the data) can be computed by (Albert, 2009; Hoff, 2009)

$$p(\mu|x) = \frac{p(x|\mu)g(\mu)}{\int_{\nu}^{\infty}p(x|\mu)g(\mu)d\mu} \quad (1)$$

We make the assumption that the prior $g(\mu)$ is equally distributed ranging from the lowest to the highest observed amount paid for a purchase. Therewith, no strong assumptions are made about the prior belief of the true distribution of $\mu$ in what follows Equation (1) becomes independent of $g$. The quantity $p(x|\mu) = \prod f(x_i)$ is called the likelihood function and describes the probability that the data follows the distribution $f(x)$. In order to resolve the posterior, Equation (1) is often computed using Markov Chain Monte Carlo simulations (MCMC). However, given the problem is only one dimensional and reasonably simplified, Equation (1) can be evaluated analytically, resulting in fast processing time of the data. Under these assumptions, the posterior simplifies to

$$p(\mu|x) = \frac{(n\bar{x})^{n-1}\mu^{-n}e^{-n\bar{x}}}{\Gamma(n-1)} \quad (2)$$

where $n$ is the number of digital receipts, $\bar{x}$ the empirical mean and $\Gamma$ the Gamma-function. For each experimental group, the posterior distribution of the mean can be evaluated and graphically compared to other experimental groups. Of particular interest is the degree to which the posterior distributions of two experimental groups overlap, as our research hypothesis are all formulated with interest in whether one experimental group spent more on average than the other one. Hence, the quantity of interest is the probability that the mean of an experimental group 1 with $\mu = \mu_1$ is larger than the mean of another experimental group 2 with $\mu = \mu_2$. Using the posteriors of Equation (1) the probability of a generic hypothesis of form $\mu_1 > \mu_2$ is given by (Hoff, 2009)

$$P[\mu_1 > \mu_2] = \int_0^{\infty} \int_0^{\mu_1} p(\mu_1|x_1)p(\mu_2|x_2)d\mu_2d\mu_1 \quad (3)$$

where $x_1, x_2$ represents the vector containing the amount spent per shopping trip of the two experimental groups that need to be compared. For non-overlapping distributions where $\mu_1 > \mu_2$ holds, the value $P[\mu_1 > \mu_2]$ converges to unity. Equation 3 can be applied to any experimental group pair to check for statistically significant differences in the mean amount spent per shopping trip. Along with this Bayesian hypothesis test, we have also carried out the Welch test as a frequents alternative.

4. Preliminary Results

The data acquired from the field study has been separated according to the experimental user group described in section 3.1. The posteriors for each group can be computed by evaluating equation 2. Already the graphical inspection of the posterior distributions among the different experimental groups gives great insight about the relative position in terms of the average amount spent per shopping trip. To further quantify the difference between the two experimental groups equation 3 can be evaluated to quantify the pair-wise probability that an experimental group spent on average more per shopping trip compared another experimental group of interest.

The graph of figure 2 shows the posteriors of pure customers (exclusive staff and operator accounts) for all experimental groups. These include users that did not receive any personalized recommendation (“NoRecomm.”), users that did receive personalized recommendations without popularity information (“NoStarsSeen”) and users that did receive personalized recommendations with popularity information (“StarsSeen”), respectively. Each posterior describes the probability distribution of the mean amount spent per shopping trip conditioned on the actual observation.
For the pure customers, a clear trend is observable where users excluded from the recommendation service spent the least amount ($\mu(\text{NoRec.})=15.6$ EUR) followed by users that received recommendations with popularity information ($\mu(\text{StarsSeen})=24.1$ EUR), and the group without popularity information with the highest average amount spent ($\mu(\text{NoStarsSeen})=31.7$ EUR). These results show first evidence supporting the earlier stated research hypothesis $H_1$ and $H_2$. By graphically inspecting figure 1 a qualitative understanding to which degree these results are statistically significant can be gained: weakly overlapping distributions are likely to have significantly different means. In the Bayesian framework, the probability of the underlying research hypothesis can be easily computed using equation 3. The probabilities of the hypothesis are $P[\mu(\text{NoStarsSeen}) > \mu(\text{StarsSeen})]$, $P[\mu(\text{StarsSeen}) > \mu(\text{NoRecomm.})]$ and $P[\mu(\text{NoStarsSeen}) > \mu(\text{NoRecomm.})]$. All probabilities are larger than 99% which indicates that the presented findings are significant. In addition, we have carried out a frequentist Welch hypothesis test to confirm the Bayesian hypothesis probabilities. The Welch test, in which the null hypothesis assumes equal means, shows for all previously mentioned combinations that the hypothesis of equal population means can be rejected with $p<4.2\%$ using a significance level of 5%.

The plot of figure 3 compares the mean amount spent per shopping trip for recommended products for male and female customers that received recommendations with and without popularity information. It is clearly observable that $\mu(\text{MaleRecomm.})=5.51$ EUR is larger than $\mu(\text{FemaleRecomm.})=3.56$ EUR with $P[\mu(\text{MaleRecomm.}) > \mu(\text{FemaleRecomm.})]>99\%$ (Welch test yields $p=1.4\%$) indicating a significant difference in the population mean and therewith supporting the research hypothesis $H_3$.

5. Discussion

Our findings have several preliminary implications: First, we had the chance to track real customer data after the introduction of a new smartphone app at a European grocery retailer. The results for $H_1$ show that personalized recommendations help to boost the amount spent on a particular purchase trip. Hence, we replicate findings about the positive effects of personalized recommendations on sales from pure online platforms. Decision aids like our smartphone application help retailers to guide customers who are faced with increasingly complex assortments. Customers’ seem to value this added service with more money spent at the retailer.

Second, we could confirm $H_2$ stating that additional popularity information (“stars”) does not amplify the positive effect of personalized recommendation. On the contrary, combining personalized recommendation with popularity information reduces customer spending. Hence, it might be advisable for retailers to avoid integrating popularity information on their smartphone apps aiming at
personalization. However, the combined recommendation type (personalized recommendation and popularity information) still outperforms no recommendation at all, reflecting a need of customers to receive guidance even in physical retail stores.

Third and with a sole focus on the effect on recommended products, we found significant differences between men and women: Men spend significantly more money on recommended products than women. This gives a first impression of important customer segmentation criteria for smartphone app design in grocery retailing. Men seem to value additional personalization-based services more than women, who seem to be more critical towards recommendations on smartphone applications.

6. Limitations and further research

Against the background of ‘research in progress’, our study has some limitations and several unanswered questions that cast doubt over the generalizability of these first results and suggest avenues for further research within this project. First, we plan to put our analyses on a more sustainable footing regarding the underlying data. At the moment, we work with data from one specific grocery retailer and consequently have to deal with a limited sample size and a potential self-selection bias (e.g., Heckman, 1979). It might be possible that only technique affine customers participate in using the app. In this regard, we already started the data acquisition process to gather more data from more app users and more retailers. This will not only pave the way to report more robust results on randomly selected samples, but also to compare results across retailers leading to valuable across-store-comparisons. In addition, based on larger sample sizes, we will be able to apply our algorithm to both overall sales receipts as well as sales receipts on recommended products only. This checks our hypotheses on more than one dependent variable and accounts for a more fine-grained view on the effects of different types of personalized recommendations with or without popularity information.

Second, our research shows some evidence how recommendations and popularity information affect sales through a behavioral route at the moment. This means that we have to observe and control for the psychological assumptions that we made within our hypotheses generation in a second step. For example, the influence on shopping experience in general is distinguished into emotional impressions that affect customers’ moods and product information that affects rational decision-making (Groeppel and Bloch, 1990; Yim et al., 2014). Consistent with the current work, recommendation agents are intrinsically focused on product information (e.g., Van der Heijden, 2006), but nevertheless could be used as a tool to impact other constructs like brand awareness or customer satisfaction. Therefore, we designed a laboratory experiment that accounts for the different field experimental conditions and sheds light on the psychological process underlying reported outcomes in this study. At this stage and based on the field data and literature available we focused on a hypothesized difference in trust mechanisms between the sexes. It is also possible though, that other important mediators are responsible for the reported differences like underlying purchase decision involvement (Shao et al., 2004), local merchant loyalty (Nobel et al., 2006) as well as hedonic or utilitarian shopping motives (Van Slyke et al., 2002) in grocery retailing. In addition, we try to integrate more segmentation criteria (e.g., age, income, etc.) besides gender in further studies, which could not be extracted out of the current field data.

Third, we examined a grocery retailer with a new smartphone recommendation app who has no online shop experience as this was the setting for our field study. The results might differ (spending on purchase trips and recommendations might even be higher), when the mobile recommendation app of the retailer had some time to establish in the market and consequently in the mind-set of customers (Hsieh and Chen, 2011). Therefore, it is worth examining firms that are already in the market or already gained some experience with smartphone recommendation apps via other platforms (e.g., via an online selling platform). These additional research steps could enable our results to be generalized to further contexts. For instance, researchers may use the chance to explore other types of mobile decision aids in IS and marketing research like location-based product information in retail settings. Besides our own next steps, we encourage researchers to continue exploration of these interesting avenues of inquiry.
References


