J. of the Acad. Mark. Sci.
DOI 10.1007/s11747-015-0441-x

ORIGINAL EMPIRICAL RESEARCH

Adaptive personalization using social networks

Tuck Siong Chung[1] · **Michel Wedel**[2] · **Roland T. Rust**[2]

Received: 21 August 2014 / Accepted: 7 April 2015

Abstract This research provides insights into the following questions regarding the effectiveness of mobile adaptive personalization systems: (1) to what extent can adaptive personalization produce a better service/product over time? (2) does adaptive personalization work better than self-customization? (3) does the use of the customer's social network result in better personalization? To answer these questions, we develop and implement an adaptive personalization system for personalizing mobile news based on recording and analyzing customers' behavior, plus information from their social network. The system learns from an individual's reading history, automatically discovers new material as a result of shared interests in the user's social network, and adapts the news feeds shown to the user. Field studies show that (1) repeatedly adapting to the customer's observed behavior improves personalization performance; (2) personalizing automatically, using a personalization algorithm, results in better performance than allowing the customer to self-customize; and (3) using the customer's social network for personalization results in further improvement. We conclude that mobile automated adaptive personalization systems that take advantage of social networks may be a promising approach to making personalization more effective.

Keywords Personalization · Social networks · News · Bayes classifier · Recommendation systems · Mobile commerce · Smart phones · Service marketing

✉ Roland T. Rust
rrust@rhsmith.umd.edu

Tuck Siong Chung
atschung@ntu.edu.sg

Michel Wedel
mwedel@rhsmith.umd.edu

1 Nanyang Business School, Nanyang Technological University, Singapore 639798, Singapore

2 Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA

The proliferation of customer data (often called "big data") has made it increasingly possible for companies to personalize their offerings (Rust and Huang 2014). All other things being equal, customers prefer offerings that are a better fit to their needs, and better customer data often give companies the information they need to serve the customer better. As a result, the literature on personalization in marketing has expanded enormously in recent years. Table 1 gives a sampling of some of the studies that have developed approaches to personalization. We contribute to the personalization literature by demonstrating that automated, adaptive personalization can work better than letting the customer customize the product herself, which questions the predominant personalization approach used on the Internet today. We further add to the personalization literature and social networks literatures in marketing by providing the first demonstration that social networks can improve personalization performance adaptively, demonstrated in the context of mobile news.

News personalization for mobile devices needs to take into account the limitations of current mobile devices. In particular, the relatively small size of the display of these devices imposes constraints on the user interface, and the bandwidth of wireless connections imposes limits on the amount of information that can be transferred. Although some prior research (e.g., Shapira et al. 2009; Liu et al. 2011) suggests that adaptive collaborative filtering can produce a better product over time in mobile devices, whether the user's social network can improve the personalization has not been demonstrated. Personalization

Published online: 01 May 2015

Table 1 Comparison of personalization approaches

Paper	Research focus	Method	Product customized / Personalized	Use of mobile devices	Adaptive approach	Use of social network
Current paper	Adaptive personalization of mobile devices	Bayes Classifier	News	Yes	Yes	Yes
Hauser et al. (2014)	Changing a website to match customer characteristics	Dynamic programming	Web pages	No	Yes	No
Urban et al. (2014)	Morphing banner advertising	Dynamic programming	Banner ads	No	Yes	No
Liu et al. (2011)	Filter and push blog articles to mobile users	Collaborative filtering	Blog articles	Yes	Yes	No
Atahan and Sarkar (2011)	Accelerated learning of user profiles	Information theory heuristics	Web links	No	Yes	No
Van Roy and Yan (2010)	Improving the robustness of collaborative filtering	Linear collaborative filtering	Movies	No	Yes	No
Shapira et al. (2009)	A personalized mobile newspaper	Collaborative filtering	News	Yes	Yes	No
Chung et al. (2009)	Adaptive personalization mobile devices	Particle filtering, model averaging	Music	Yes	Yes	No
Hauser et al. (2009)	Matching look and feel of website to user	Dynamic programming	Broadband subscriptions / Web pages	No	Yes	No
Liang et al. (2008)	Comparison of semantic-expansion to keyword approach	Semantic-expansion	Digital documents	No	Yes	No
Moon et al. (2008)	Predicting product purchase using customer similarity	Spatial choice modeling	Insurance policies	No	No	Yes
Das et al. (2007)	Personalization of Google News	Collaborative filtering	News	No	Yes	No
Ying et al. (2006)	Accounting for missing ratings to improve recommend-ations	Mixture and Hierarchical Bayes	Movies	No	No	No
Zhang and Krishnamurthi (2004)	Customizing online promotions	Purchase incidence and brand choice model	Online promotions	No	No	No
Ansari and Mela (2003)	Customizing e-mail to increase Web site traffic	Mixture of Dirichlet process probit model	Emails	No	No	No
Ansari et al. (2000)	Design of a recommendation model	Hierarchical Bayes	Movies	No	No	No

benefits from the use of social networks have been shown by Moon et al. (2008), however their approach is neither adaptive nor demonstrated in mobile devices. Our study thus contributes in several important ways to the personalization literature.

One promising approach to personalization is to adapt ("morph") the product based on what is observed about the customer's behavior over time. Systems that implement this approach are known as "adaptive personalization systems" (Chung et al. 2009) or "morphing" systems (Hauser et al. 2009; Hauser et al. 2014; Urban et al. 2014).

The adaptive personalization approach has several key characteristics:

1. It is done automatically using algorithms.
2. It requires no proactive effort on the part of the customer.
3. It observes customer behavior and adapts the product over time.

With more and more of business revolving around information services, the cost of changing products over time is less and less of an obstacle, meaning that the key is instead determining how the products should change. Adaptive personalization can be accomplished at low per use cost by using automated algorithms, because customer behavior data with respect to information products can be collected easily. Removing any customization responsibility from the customer also reduces the customer's burden in terms of time or effort.

We posit that (1) adaptive personalization will produce better outcomes over time as it learns more about the customer, (2) adaptive personalization will work better than self-customization, and (3) incorporating information from the customer's social network will further improve performance. Although we cannot test whether the above is true with every conceivable adaptive personalization system, we are able to test the above on a particular example—an adaptive personalization system for mobile news. We develop and implement the system, using observed data from the customer's own behavior, plus information from their social network, and test the benefits of the system against various alternatives.

Implementation: adaptive personalization of mobile news

Mobile devices have rapidly become increasingly powerful, versatile and ubiquitous. Smart phones and similar devices are now blurring the line between phones and personal computers, having global positioning, web-browsing, and social networking capabilities. Global mobile devices reached seven billion in 2013 and by 2016 mobile app revenues are projected to hit $46 billion (Karr 2014). For more and more people, mobile devices are now a main source for telecommunication, social networking, shopping, entertainment, and (the focus of the present study) news. In the United States, 79% of the shoppers use their smartphone for browsing and shopping on website and apps (Karr 2014). Historically, advances in technology have enabled leading firms to adapt and personalize products/services to better fit the individual (Varki and Rust 1998; Khan et al. 2009). Mobile devices present many opportunities for such personalization (Lyytinen and Yoo 2002). Mobile devices have overtaken desktop computers and notebooks as the preferred medium to receive, transmit and consume information. They are more integrated into an individual's personal life and represent a more natural way in which a consumer consumes digital services (e.g. mobile news) in different contexts. But, as yet, there is only limited research showing how to personalize in a mobile environment (Chung et al. 2009; Shapira et al. 2009), and what the benefits are of doing so. This is even more important because compared to stationary online access (Ansari et al. 2000; Ying et al. 2006), mobile devices are hampered by consumers' shorter attention spans and more volatile preferences, concerns over privacy, and limits on bandwidth, computation, and display.

Further, the rapidly growing literature on social contagion has shown that social proximity to others can influence whether consumers try new goods and services (Bell and Song 2007; Godes and Mayzlin 2009; Iyengar et al. 2011). Increasingly consumers access their social network via mobile devices, and it has been argued that therefore social networks are a promising path to personalization. Indeed, consumers actively share product information and recommendations via their social networks. The extent to which social networks provide a promising avenue for adaptive personalization, however, is not known.

Personalization of mobile news

The study of news personalization is important in its own right. News has traditionally been mass-produced, with television, magazines and newspapers all catering to mass audiences. The US newspaper industry accrues over ten billion dollars annually in circulation revenue and over twenty billion dollars in advertising revenue (Newspaper Association of America 2014). For the past several years the advertising revenue in print has been falling and the advertising revenues from online and mobile news have grown to over three billion dollars annually (Edmonds et al. 2013). Over half of the American public refers to the Internet as their main source of national and international news. About 20% of Americans reports to have seen news on a social network on a daily basis (Caumont 2013). The growth in online news consumption is driven by the growth of net books, smart phones, e-readers and tablet computers, which allow users to subscribe to many newspapers at a fraction of the cost of the printed versions. About one in four consumers in the US consumes news through mobile devices, and the introduction of smart phones has increased traffic to news sites by about ten percent (Subramanian 2012).

Next to cost reduction, the success of online and mobile news services is driven by their appeal to smaller and more targeted audiences. Many of the largest news websites, including Yahoo News, MSNBC Digital Network, AOL News, CNN, and NY Times, therefore enable personalization of news content. Mobile phone apps of the major news providers, as well as popular mobile news services, such as Pulse and Flipboard, all offer personalization options. The vast majority, however, only allow the user to personalize their news manually (self-customize) on the basis of explicit selection or evaluation of news categories by users.

One notable exception (Das et al. 2007; Liu et al. 2010) personalizes Google News based on the user's reading history. Another develops a personalized e-newspaper (Shapira et al. 2009), aggregated from various online sources, that adapts over time to the reading behavior of the user. Current commercial efforts along these lines include Drippler (TechCrunch 2014) and News360 (Google 2014).

Online consumers spend only about 8 minutes a day reading news, compared to about 30–50 minutes for readers of print (Kirchoff 2010). Mobile users may have even shorter attention spans. Thus, without personalizing news content, there simply is not time for a mobile user to find the news he/she seeks. This makes personalization for mobile devices essential. We therefore investigate the key issues surrounding adaptive personalization by developing and implementing a new system that changes the way news can be consumed through mobile devices by delivering highly personalized real-time pod-casting of news. The system personalizes news items in real time by filtering news feeds based on past reading behavior, and automatically delivers news on a mobile device. Most importantly, the system automatically discovers new material as a result of shared interests in the user's social network, without "friends" in the network actively having to share news items of joint interest. The system improves the personalization process through an iterative feedback loop, resulting in a cycle of personalization (Adomavicius and Tuzhilin 2005).

The interface of the system is simple and intuitive, and takes into consideration the constraints of the mobile devices. It works through a mobile text-scroll that scrolls news headlines. The user may decide to halt the scroll through the touch-screen, read the headline, and bring up the underlying full text and/or audio or video clips. Rather than indiscriminately scrolling all news feeds, the feeds are filtered using a user-specific text-classification system.

We use a modified Naïve Bayesian algorithm to implement our adaptive personalization system.[1] Naïve Bayes classifiers consistently perform as well or better than most of the other, often more advanced, techniques that have been proposed for personalization (Hand and Yu 2001). The technical details of this classification algorithm are given in Appendix 1. Figure 1 shows the logic behind the adaptive personalization system and how it employs the classification algorithm. The algorithm is adaptive and thus captures individuals' change in reading preference, e.g. when individuals begin to seek variety in the articles they read. We implemented the algorithm on HP PDA's, but the algorithm can easily be applied to other mobile devices such as smartphones. In our illustration, the mobile device connects to an RSS (Really Simple Syndication) website, which has assembled a large number of news articles. The news articles are downloaded into the mobile device and the words in the articles are catalogued.

Based on what the customer has read before, keywords are chosen that are most predictive of whether the customer would read an article. That is, words that appear in articles that are read, but don't appear in articles that are not read, are maximally predictive. The Bayesian algorithm then estimates the probability that the customer would read each article, if that article were presented to him/her. Based on these probabilities a personalized list of news articles is constructed and their headlines are presented in a news-scroll on the mobile device. The customer can choose to read an article by tapping on the headline in question, which brings up the article. Whether the customer chooses particular articles or not then becomes additional information that is used in the subsequent rounds to further improve the Bayesian estimation, and provide ever better article lists for display on the news scroll. The system thus allows for continuous updating of the classification of news at the individual level, without any proactive input from the user. Based on the design of the system, which learns the customer's preferences over time, we predict:

H1: The adaptive personalization system will improve its ability to select articles that will be read more fully, as the system records and analyzes an increasing amount of the reader's behavior and preferences.

Information overload and self-customization

RSS (Really Simple Syndication) is a format for distributing and gathering news content from web sources, including online newspapers and magazines. RSS creates news feeds that allow a consumer to see when websites have added new news content. It gets the latest headlines, summaries, and links to full text, audio- and video-clips, without the user having to visit the websites the feeds are taken from. News aggregators are software solutions that automatically retrieve RSS news feeds and compile links to content from electronic news sources. Examples are NewzCrawler, NewsGator, Yahoo and AOL. Mobile applications of RSS aggregators include those on NewsGator, Pulse and Flipboard. They make it possible to track and manage RSS content on mobile devices, especially tablets, e-readers, Blackberries, PDAs and Smart-phones.

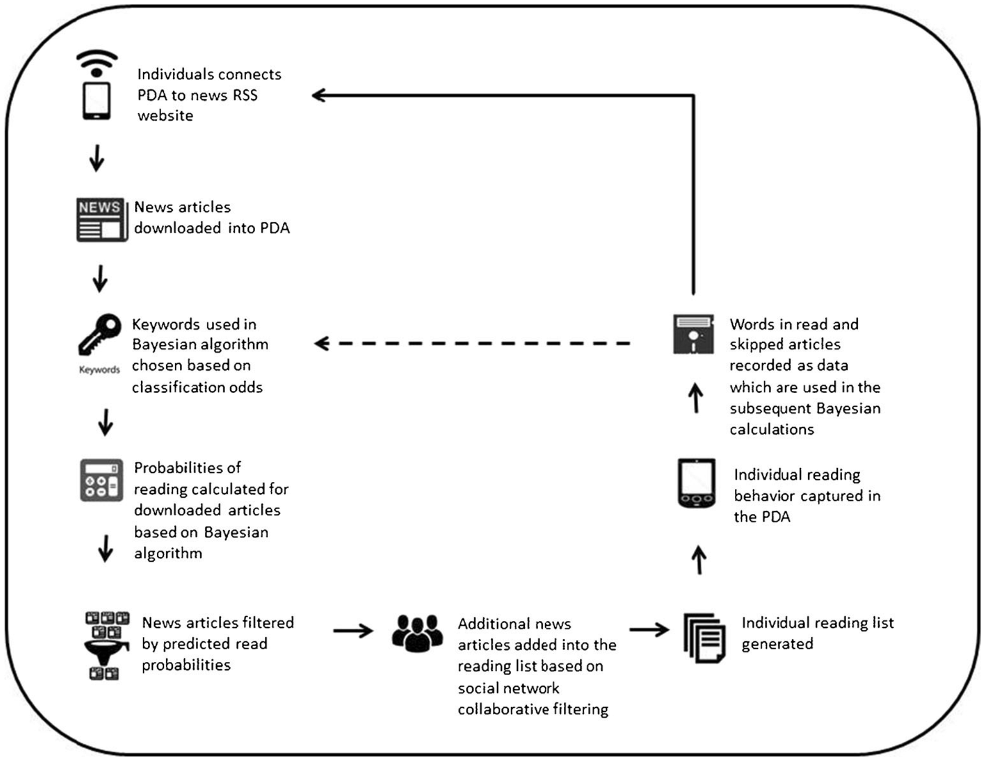
However, subscribing to RSS feeds may result in a substantial information overload for users. Feeds relevant to a specific user easily get lost in the streams of feeds of subject areas of little interest. Therefore, most news sites offer personalization options, in which they allow the user to personalize their news manually (self-customize) by selecting relevant news categories or news sources. However, although this approach is widely used, some research has cast initial doubt on customers' ability to perform such self-customization effectively (Franke et al. 2009).

Moreover, news by its very nature is dynamic and non-domain specific. A single news story can cover topics ranging from sports sponsoring and entertainment, to geographic regions and commercial organizations at the same time. This negatively impacts the effectiveness of personalization of news based on self-customization, and these classifications may therefore still result in hundreds of feeds that people don't read. The adaptive personalization system proposed here addresses that problem through automated selective aggregation of relevant news content for individual users, without requiring them to provide explicit input in the system. We therefore expect that the flexibility provided by personalizing according to specific words rather than pre-selected categories will produce better performance. We predict:

H2: The adaptive personalization system, based on dynamically updated predictive keywords, will out-perform self-customization in terms of serving articles which will be more likely to be read and sorting out articles which will be less likely to be read.

[1] We also tested a Bayesian Logistic Regression model. That model, although seemingly more sophisticated than the Naïve Bayes approach, and a better performer in in-sample testing, actually performed substantially worse in out-of-sample tests, suggesting that the added complexity of the Bayesian Logistic Regression model resulted in over-fitting. Thus, we focus on the Naïve Bayes algorithm for the remainder of this paper.

Fig. 1 An adaptive personalization system for mobile news



Using social networks to improve personalization

The literature on social networks has extensively documented the influence of peers on the adoption and usage of products and services (Bell and Song 2007; Hartmann et al. 2008; Godes and Mayzlin 2009; Hartmann 2010; Nair et al. 2010; Iyengar et al. 2011; Narayan et al. 2011). Peers, as opposed to experts, exert greater influence on choices for products that involve personal taste (e.g. fashion), compared to products that are more technical in nature (Wang et al. 2013). Personalization using social networks might be more effective because consumers rely on choices of their peers as an additional attribute (Bell and Song 2007), or revise attribute preferences as a result of peer influences (Narayan et al. 2011). The importance of social networks is evident: Facebook has over 400 million mobile users, and the two largest mobile social networking services, Foursquare and Instagram, each have about 15 million users (Keath 2011). Further, Facebook has partnered with the Wall Street Journal and the Washington Post, amongst others, to release mobile Social Reader apps. Social news aggregators Digg.com and Reddit.com allow users to submit stories and/or moderate them. Many online news sites, including the NY Times, Yahoo and AOL, allow readers to click a button to share news articles with friends in their networks on Facebook, Linkedin and Twitter. All these developments signal major potential to integrate news consumption better into people's daily lives (Lyytinen et al. 2004).

Still, in spite of these developments, only about 9% of consumers click on news links on social network sites (Subramanian 2012). For social news filtering to work, users are currently required to proactively share news-items of interest. Not only may users be reluctant to spend the effort required, but, especially as the size of individual social networks continues to grow, issues of privacy may prohibit users from disclosing articles that expose their opinions and interests. Privacy and lack of trust are potential obstacles in the adoption of electronic information services, which use detailed data about consumers to personalize transactions (Fichman and Cronin 2003). Therefore, the adaptive personalization system that we develop and implement is based on automated, unobtrusive and anonymous sharing of content across the social network, without the user being aware of the precise source.

Importantly, when social influence occurs through observational learning (Zhang 2010), the influence from peers has been shown to be strongest if an individual is uncertain of her preferences (Narayan et al. 2011). We thus employ the user's social network for cases in which a user's preference for a news article is ambiguous, because for these articles uncertainty is the highest and the potential for peer influence the largest. Therefore, the proposed system not only learns the users' preferences from her news consumption, but, moreover, allows for automated and unobtrusive sharing of news items of potential interest in the network.

The proposed system thus utilizes an individual's social network to improve on the personalization of news article on mobile devices. Incorporating additional information into a personalization in the form of peers' preferences helps to tackle the cold-start problem presented in most systems and increases prediction reliability. Even if the predictions of a personalization system are reliable, shared preferences may still help increase the likelihood that an individual is likely to accept the personalization. The motivation for this stems from the fact that individuals in the social network groups are more homogeneous. This generally arises due to either similarity (homophily) or social influence (induction). Homophily is a form of social selection in which people who are similar tend to attract each other and form social groups (Mcpherson et al. 2001). Induction, or social influence, involves an individual's attitudes or decisions being influenced by social contact with peers (Cialdini 2001). Within the growing literature on social network effects, for example, Ansari et al. (2011) utilize connectivity in a social network (homophily) to improve the marketing of products, and Iyengar et al. (2011) use social contagion (induction) to explain product diffusion.

This research investigates the impact of the social network on news consumption through both homophily and induction. The makeup of the social network consists of peers in which individuals have a high frequency of contact and close relationship ties. Strength of a relationship affects the likelihood that an individual is affected by the recommendation of the peers (Levin and Cross 2004). Thus, popularity of a news article within a social network is likely to be more influential than the popularity of a news article among the general public. Trust of peers in the social network may translate into trust of the personalization system. This is important, because empirical findings suggest that individuals are more likely to accept the personalization of a system they trust (Smith et al. 2005).

Even though individuals in the network are homogeneous, they do not have identical news reading preferences. Individual interests may also change with time. Another potential advantage of incorporating the social network into the news recommendation has to do with serendipity. It may offer the important advantage of preventing the algorithm from zooming in on a too narrow set of news articles early on, by introducing a certain level of 'surprise'. Based on the above arguments, we expect that use of the social network in personalization will improve performance of the system, due to social influence (induction), similarity (homophily) and the enablement of serendipity. We predict that:

H3a: Incorporating induction and homophily into adaptive personalization will improve the system's ability to serve articles which are read.

H3b: Adaptive personalization that takes into account social network effects will out-perform personalization based on general article popularity in terms of serving articles which are more likely to be read and sorting out articles which are less likely to be read.

Contributions of the proposed adaptive personalization system

The proposed adaptive personalization system provides users with several benefits. The system automates frequent decisions that many consumers prefer to avoid making. Removing the need for making a decision about which news articles to read reduces search effort (Bechwati and Xia 2003) and the cost of processing information (Shugan 1980), which makes exploring and reading news more pleasurable and satisfactory (Häubl and Trifts 2000; Bechwati and Xia 2003). In addition, the system does not require asking individuals explicitly for their reading preferences or news categories of interest. This avoids problems of missing information due to users' unwillingness to actively provide their evaluations of news articles (Ying et al. 2006), of unfamiliar items receiving less positive evaluations (Cooke et al. 2002), and of users idiosyncratically responding to rating scales provided to them (Rossi et al. 2001). Further, the proposed adaptive personalization system assists users in unobtrusively discovering relevant news through their social network, without the need to actively share content, and without an undue influence of heavy users or opinion leaders (Godes 2011). Finally, the algorithms used are both robust and scalable. While the computations required for personalization can be run either on the mobile device or on a central server (cloud computing), scalability presents an important challenge for personalization systems regardless of whether they use ratings input or not. Therefore, we believe that the proposed system presents an opportunity to investigate the key questions on the potential benefits of adaptive peronalization.

Outline of the system

A prototype of the system was developed and implemented on Hewlett Packard IPAQ212 PDAs and programmed in Visual Basic .NET Compact Framework. Its key features are:

1. Automated Standard RSS news feed input,
2. Mobile filtered RSS news-feed text scroll with a Stop/Explore interface,
3. Individual-specific text classifiers calibrated on the content of past articles,
4. RSS news feed input based on the individual's social network.

System design

Extraction of news headlines An RSS feed reader was developed for the mobile device, which filters the RSS feeds from a news website. News filters are personalized for each individual user and the filtered news is downloaded into the mobile device.

The user interface Sample screen-captures from the PDA's user interface are shown in Fig. 2. For our purposes, the user interface was kept as simple as possible, without limiting the functionality or the collection of users' reading behaviors. Feature fatigue (i.e. frustration with the multiple features on a device) may creep in when the user interface gets too complex (Thompson et al. 2005).

The software starts up with the main screen showing the option to "Display Available Headlines". When this option is selected, the news headlines of the filtered RSS feeds are scrolled one headline at a time. A batch of 15 to 30 news feeds is scrolled at any given time. The scroll speed of an individual headline can be adjusted, with a lag of 1 to 9 s between headlines. Popularity of news articles, that is, the number of peers in the social network that have read the article, is displayed in the headline. If the "*Read News Story*" button is not tapped when a particular headline is shown, the software records that this headline is skipped. When the user taps on "*Read News Story*", the next screen with the caption: "*News Content*" pops up. On this screen the content of the article is displayed. When the content spans more than a full screen, the reader can use the scroll bar on the right to access the later portion of the article. When the "*Back*" button is tapped, the software returns to the main screen and the scrolling of the headlines continues.

Data capture The data captured from the mobile device (stored in the reading log) includes:

1. Whether a reader skips the headline without choosing to read the article in detail, or taps on the "*Read Story*" button to read the article,
2. The time that a reader spent reading an article: the time between tapping the "*Read Story*" button and the "*Return*" button (seconds),
3. The actual content of the news stories (dummy variables indicating each word in the text).

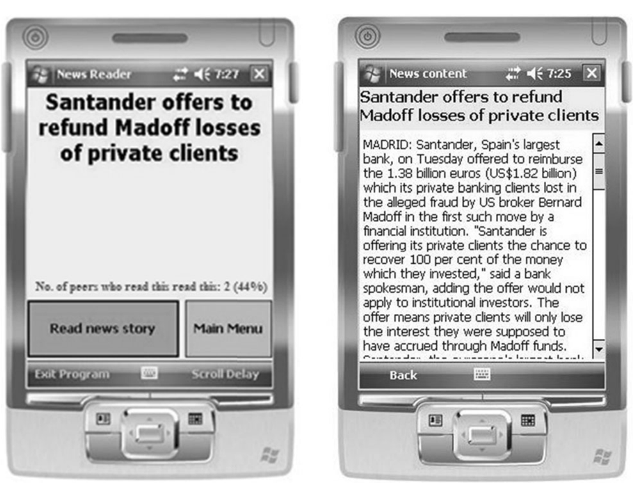


Fig. 2 Screen captures from the mobile device

Field studies of adaptive mobile personalization

We conducted two field studies: the first (Study 1) collected data on participants' reading behavior on mobile devices to compare how well reading behaviors are predicted ex-post by the proposed algorithm and two other benchmarks. The second study (Study 2) implemented the adaptive personalization system in a field study and reports reading behaviors resulting from personalization. Figure 3 provides a high-level overview of the two studies.

Study 1: Comparison with benchmark procedures

Participants and procedures We compared the proposed algorithm described in Appendix 1, against two benchmarks that allow for investigation of our questions on the effectiveness of adaptive personalization, on a dataset we collected using a total of 48 participants from a major university. The number of subjects used is necessarily small, due to the intensive nature of the data collection, involving the customized programming and actual field use of mobile devices, however the data collected include close to 7000 data points for our analysis, and the studies do present a real-life implementation. Participation was voluntary and participants were not selected using any specific characteristics, but they were paid and needed to be able to access a wireless network to download news articles. Participants were told that the purpose of the study was to validate a news recommendation system that customizes news articles. The selection criteria included that participants needed to be members of an existing Facebook group. Since the participants were all students from the same university, the Facebook groups shared a common background and possibly similar reading preferences.

The algorithms for adaptive personalization of news described above were implemented on Hewlett Packard IPAQ212 PDA units, which were then used by participants to download and display the news articles, and to collect data on users' reading behaviors. Participants were issued a mobile device to download and read mobile news at their own time and pace and were briefed on how to operate the PDA and how to use the news personalization software. Data collection lasted from 7 to 10 days per participant giving an average number of cycles of 4.9 (a minimum of 2 and a maximum of 8). The participants used the mobile device as they went about their daily life, and while carrying out their daily activities in 'normal situations' (i.e. they were not in a laboratory when they interacted with the mobile device). They were assured that there were no right or wrong news articles to read,

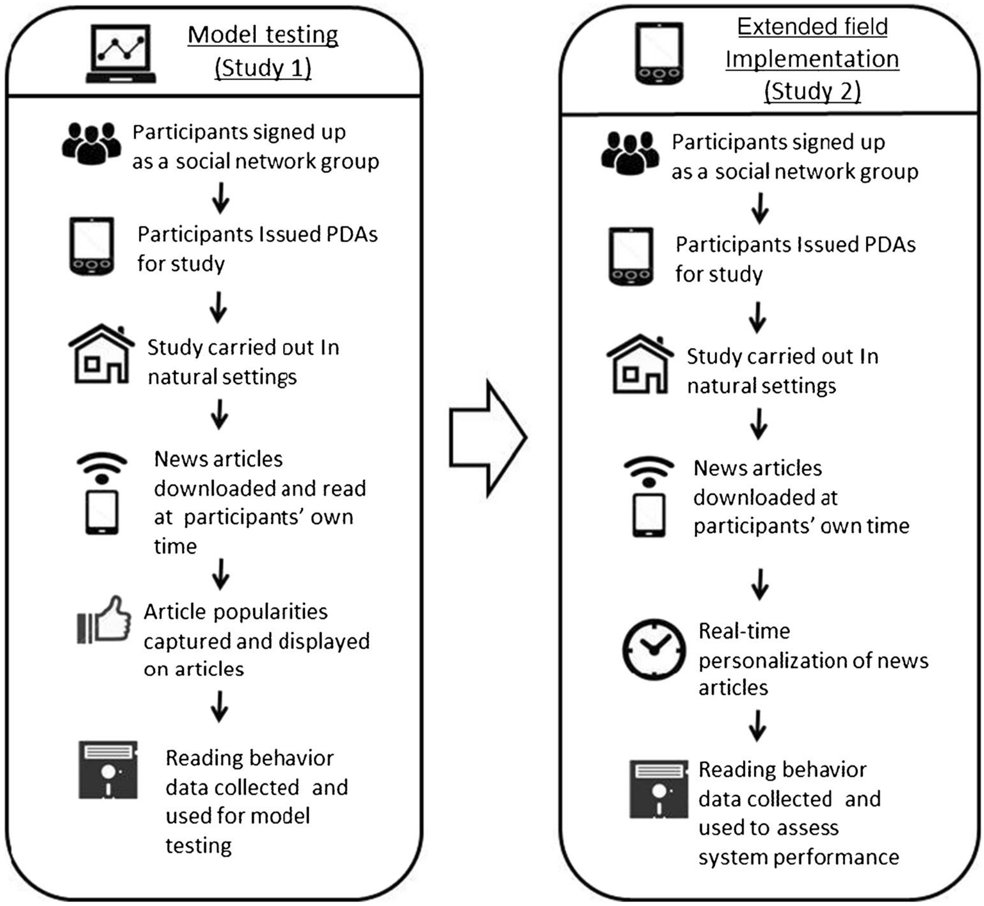


Fig. 3 Overview of the research

and were informed that their reading behaviors would be recorded with the purpose of downloading new news articles that matched their preferences better.

During the downloading process, RSS news feeds were extracted from a popular business news site and delivered on the mobile devices. Even though the focus of the news site is on business news, it provides news in other general categories of world, regional, local, and sports news allowing it to capture diverse reading interests of our study population. Importantly, in this benchmarking study, the news articles presented to participants were not personalized. Therefore, participants who were in the same randomly assigned sub-group were fed the same collection of news articles. This made it possible, after the reading data are collected, to assess how well the different algorithms and procedures predicted reading behavior ex-post. The reason for not personalizing the news feeds to individual participants in this phase of the research was that if articles are personalized, the user obviously would not have access to articles that are not recommended. This would create a strong selection effect that would bias the results towards the specific system that is used to personalize news and the specific news items downloaded. This would make it impossible to compute “true negatives” (articles that are not downloaded, and not read), which are needed to evaluate the relative performance of the systems. By using one common dataset involving users exposed to the same articles the results are not biased towards any particular model, and allows us to measure true and false negative rates.

We used two seconds as a threshold before considering an article to be truly read in the benchmark study. A user may click on an article and figure out that s/he is not interested after a only few seconds and go back to the main screen. If so, such articles would be incorrectly classified as being of interest to the user. In the reading-time data a spike occurs where users read the article for two seconds or less, which amounts to about 8% of the articles that are read. We reason that 2 s is enough for an individual who is skimming to decide if an article is of interest or not.

The 48 participants in the benchmarking study come from three separate Facebook groups. The three social network groups range from 15 to 18 individuals (larger real-life networks may produce results that are stronger and closer to real- world settings). We split the members of each social network group randomly into two sub-groups, A and B. Sub-group A had the opportunity to read half of today’s articles and half of yesterday’s articles, the latter being previously presented to sub-group B. This allowed us to capture the popularity of yesterday’s articles, based on how much they were read among the sub-group B members. We presented the information on article popularity on yesterday's articles to sub-group A members. The presentation of articles to sub-group B was similar but reversed. During the debriefing of the experiment, participants were informed about how article popularity information was collected and displayed.

Reading histories were automatically shared in the background to facilitate the social filtering aspect of our approach. The files were kept in a web-based file sharing site. The reading histories of the individuals in a particular network were then automatically downloaded into the target individual's PDA, which made the file sharing "anonymous" within the social networks. We did not disseminate actual reading behavior of individual peers in the social network to protect each individual user's privacy. Thus, the design of the system ensured that users were not informed about individual peers' reading histories, but only on the number of peers who read the articles. Approximately half of the dataset collected consisted of articles which were flagged with popularity information, and half did not. Analyzing the performance of our algorithms separately for these flagged and not flagged articles allowed us to tease out the effect of homophily versus induction.

Performance measures and results We evaluate the performance of the algorithms using a standard performance measure from the information retrieval literature (Van Rijsbergen 1979; Sebastiani 2002). Given a test set of articles, a two-by-two contingency table of the counts of true positives (articles predicted to be preferred and actually read; TP), false positives (articles predicted to be preferred but not actually read; FP), true negatives (articles predicted to be not preferred and not actually read; TN) and false negatives (articles predicted to be not preferred but actually read; FN), can be constructed. Based on this table, two commonly used measures of classification effectiveness are called Precision $= \frac{TP}{TP+FP}$, and Recall $= \frac{TP}{TP+FN}$.[2] In the context of this paper, precision is the proportion of personalized articles which is read, truly reflecting an individual's preference. A higher precision results from an individual reading a larger number of articles which the system predicted would be preferred. Recall is the proportion of truly preferred articles in the system which is personalized to an individual. A higher recall results from the system missing out on fewer articles which an individual would read, but were not recommended.

There is an inherent tradeoff for a personalization system between precision and recall; high recall can be achieved at the price of very low precision. To provide a more balanced assessment of a system's performance, the F1 measure is used. The practical significance of F1 is that an individual would like a system that has both high precision, thus not personalizing articles that he/she don't want, while having high recall, that is not missing out on articles that he/she would like to read. The measure $F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$, combines precision and recall in a single quantity as a weighted average (Cohen and Singer 1999). F1 is bound between 0 and 1, and can be interpreted as a probability: the closer it is to 1 the better. We also computed two additional measures, namely Accuracy, which is calculated as: $(TP+TN)/(TP+FP+TN+FN)$, and Specificity, given by: $TN/(TN+FP)$. Accuracy is the proportion of both true positives personalized and the proportion of true negatives not personalized by the system. A higher accuracy indicates that the system is better able to predict read and unread articles. The problem with accuracy is that if there are many true negatives, accuracy can be artificially inflated simply by not serving a very large proportion of the articles. In the case where an individual only prefers 10% of the articles for example, an accuracy of 90% is achieved by not serving any articles at all. We thus introduce another metric that we call "positive accuracy", which measures the number of true positives personalized as a proportion of all the articles available in the system. Positive accuracy is defined as $(TP)/(TP+FP+TN+FN)$, that is the same as accuracy, but with the true negatives removed from the numerator. Specificity is the reverse of precision: it is the proportion of articles which are not preferred that are actually not personalized. In order words, it represents how well the system sieves out unwanted articles. For easy reference, the definitions and formulas for all of the classification measures are provided in Table 2. To give us an indication of expected norms for reasonable system performance, we refer to Li et al. (2011) who evaluate the performance of recommendation systems involving different recommendation methodologies. Their results give a range of F1 values from 0.328 to 0.502, precision values from 61 to 83% and recall values from 9 to 56%.

The performance measures used, including precision and F1, have too complex a form to permit a traditional test of significance. Precision and F1 measures are nonlinear and their sum of differences may not have a Normal distribution. To overcome this we test for the significance differences in the performance of adaptive personalization systems using a (computationally intensive) randomization test (Lunneborg 2000). The test is explained in Appendix 2. For our study we use 10,000 randomizations in each test and a 5% threshold to reject the null hypothesis.

Table 3 shows the predictive performance of the Adaptive Personalization (AP) model. In other words, the table shows how well the estimated parameters previously calculated on the calibration data predict individual preferences for a new set of articles. Different personalization models are compared, including a model that personalizes based on explicit category preferences, a model based on popularity, and models with and without the social network component (similarity and/or

[2] Although the use of these terms may seem non-standard to a marketing audience, we retain them to maintain consistency with the classification literature.

Table 2 Performance measures

		Actual reading behaviour	
Prediction by the personalization system		Read	Not read
	Preferred	True Positive (TP)	False Positive (FP)
	Not preferred	False Negative (FN)	True Negative (TN)
Measure	Formula	Definition	
Precision	$\frac{TP}{TP+FP}$	Proportion of personalized articles which is read, truly reflecting an individual's preference	
Recall	$\frac{TP}{TP+FN}$	Proportion of truly preferred articles in the system which is personalized to an individual	
F1	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Weighted average of precision and recall allowing a more balanced assessment of a system's performance	
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$	Proportion of both true positives personalized and the proportion of true negatives not personalized by the system	
Specificity	$\frac{TN}{TN+FP}$	Proportion of articles which are not preferred that are actually not personalized	
Positive accuracy	$\frac{TP}{TP+FP+TN+FN}$	Number of true positives personalized as a proportion of all the articles available in the system.	

social influence). Comparing these different models allows us to provide an answer to the questions outlined above.

The basic AP model (method 1) uses only the reading odds calculated for each article to predict preferences. It does not incorporate the social network. With a specificity of .958 this basic model does extremely well in terms of screening out articles that individuals do not want. The high specificity together with an accuracy of .783 means that the model is highly effective in personalizing articles that individuals prefer. This fulfills the objective of a personalization system in providing targeted offerings that are a better fit to individual needs. A high accuracy and specificity is invaluable for managing the problem of information overload and shorter attention spans. These results and the results in Study 2 support H1. However, with a positive accuracy of .057, this model provides an extremely targeted set of news articles. It has the potential problem of zooming in too narrowly on a set of news articles early on and not introducing surprises into the personalization to allow for and adapt to taste changes.

To improve on the personalization, we look at the performance of a model that incorporates social influence (induction) (model 2), a model that incorporates similarity (homophily) (model 3) and a model that incorporates both social influence (induction) and similarity (homophily) (model 4). Incorporating the social network enables individuals to discover new material as a result of shared interests, utilize peers preferences as an additional decision criterion and mitigate a possible cold-start problem. Incorporating the social network, the model's positive accuracy value grows by 93.0% in model 2, 94.7% in model 3 and 188.8% in model 4. This is accompanied by an increase in recall values to .456, .457 and .679 respectively. The higher positive accuracy and recall values indicate that the proportions of truly preferred articles increase, which supports H3a. One point to note is that incorporating the social network effect increases only the positive personalization while somewhat decreasing the ability of the system to screen out articles. This explains why the precision drops from .639 in model 1 to .416 in model 2, .349 in model 3 and .333 in model

Table 3 Predictive performance of the classification algorithms

No.	Method	Positive accuracy	Precision	Recall	F_1	Accuracy	Specificity	True positives (TP)	False positives (FP)	True negatives (TN)	False negatives (FN)	Total articles (N)
1	AP (No induction nor homophily)	.057	.639	.234	.343	.783	.958	392	221	5026	1283	6922
2	AP (Induction)	.110	.416	.456	.435	.713	.795	764	1074	4173	911	6922
3	AP(Homophily)	.111	.349	.457	.395	.662	.728	766	1429	3816	911	6922
4	AP (Induction and homophily)	.164	.333	.679	.446	.593	.565	1137	2282	2965	538	6922
5	Article popularity	.160	.255	.663	.369	.451	.383	1108	3237	2013	564	6922
6	Category preference	.115	.309	.475	.374	.616	.661	796	1781	3466	879	6922

4. There is a tradeoff between precision and recall, and the F1 score helps to determine if a drop in precision is more than compensated by the increase in recall. This is the case when we move from model 1 to model 4, where the F1 value improves from .343 to .446. This difference in F1 measure between model 1 and 4 is highly significant ($p<.001$).

Next, an exploratory analysis was conducted of the percentage of participants who read an article as a function of its popularity (models 2–4). The results show that this relationship is significant (F-value=26.768; $p<.001$). News articles with a popularity of 0 are read by 17.0% of participants, those with a popularity of 1 by 27.2%, those with a popularity of 2 by 32.3%, with a popularity of 3 by 51.6%, and so on, while finally news articles with a popularity of 6 are read by 57.1% of participants. Thus, from this exploratory analysis it seems that the effect of induction, or social contagion, is strong.

The F1 values are .435 for model 2, a model that incorporates social influence/induction. This value is higher than that of .395 (sign $p<.001$) for model 3, the model that incorporates similarity/homophily only. This demonstrates that social influence (induction) dominates in the acceptance of social recommendations, because the performance for model 2 is better than that of model 3 in almost every aspect. We recognize that articles flagged with popularity may be higher quality articles with wider appeal. However, this should not cause the difference in the results between models based on social influence and similarity. Higher quality articles should have a higher probability of being read in both cases.

Given the strong influence of article popularity on the acceptance of recommendations, we also compare our algorithm to an algorithm that predicts news article preference based on popularity only. In this "popularity model" (model 5), an article is predicted to be preferred by an individual in sub-group A whenever one individual in sub-group B (from the same social network) read it the day before. The results in Table 2 show that model 2, the model that incorporates social influence from the individual's own social network outperforms this approach on precision (sign $p<.001$), accuracy (sign $p<.001$), specificity (sign $p<.001$), and most of all, the F1 measure (sign $p<.001$). These results support H3b. In model 2, we use readership in the individuals' social network to "break ties" in cases in which the system is not sure whether the article would be read or not. In model 5, the calculated odds of reading are ignored for popular articles and all popular articles in the social network are served to individuals. This explains the difference in accuracy and specificity between model 2 and 5. Model 2 is better able to account for unique individual article preferences. Therefore, recommending articles purely using the popularity of the article among the general public may not be an effective approach. From the present results, it appears that the performance of this approach may suffer greatly, as it does not account for individual differences in taste.

Finally, we compare our AP algorithm with the current industry practice of presenting articles to individuals on the basis of users' stated preferred news categories (self-customization) (Model 6). In this study, participants were asked to rank news categories. The algorithm predicts that articles in the top two ranking news categories for each participant will be read. The performance of model 6 is aided by explicitly asking individuals for their preferences. This circumvents the cold-start problem and increases its performance, especially for accuracy and specificity. Despite this, the proposed adaptive personalization algorithm (model 4) performs better in terms of positive accuracy (sign $p<.001$), precision (sign $p<.005$), recall (sign $p<.001$) and the F1 measure (sign $p<.001$), as compared to model 6. This supports H2. Participants are much more selective than what their category preference indicates, which negatively affects both the precision and recall of the self-customization algorithm. This casts some doubt on the current industry practice of using self-stated news category preferences as a basis for personalization, and shows that the adaptive personalization system works better.

Figure 4 provides an overview of the findings from Study 1. The results of this benchmarking study thus indicate that the AP algorithm generally has the highest predictive performance, especially when the social network component is included. This study shows that in our study personalizing automatically, using a personalization algorithm, results in better performance than allowing the customer to self-customize, and that using the customer's social network for personalization results in further improvements. To investigate whether the adaptive personalization algorithm improves personalization performance over time, we illustrate the implementation of the AP algorithm in a second field study.

Study 2: Extended field implementation

Because in the benchmarking study (Study 1) the news articles were not personalized over multiple cycles, we implemented the AP algorithm in a second field study (Study 2) and investigated the effects of personalizing news across a number of cycles in which the personalization algorithm adapts to user tastes better and better. Data were collected from 109 participants, providing more than 9000 usable data points. Participants were undergraduate students from a major university and signed up for the field study in groups of three, which constituted their a-priori defined social network. While the size of the social networks is smaller than in the benchmark study and in many real-life applications, the purpose here is to illustrate the algorithm, while avoiding the influence of endogeneous network formation and minimizing the influence of similarity (homophily). Hewlett Packard IPAQ212 PDA units were used to download and display the news articles, and to collect data on users' reading behaviors. The adaptive

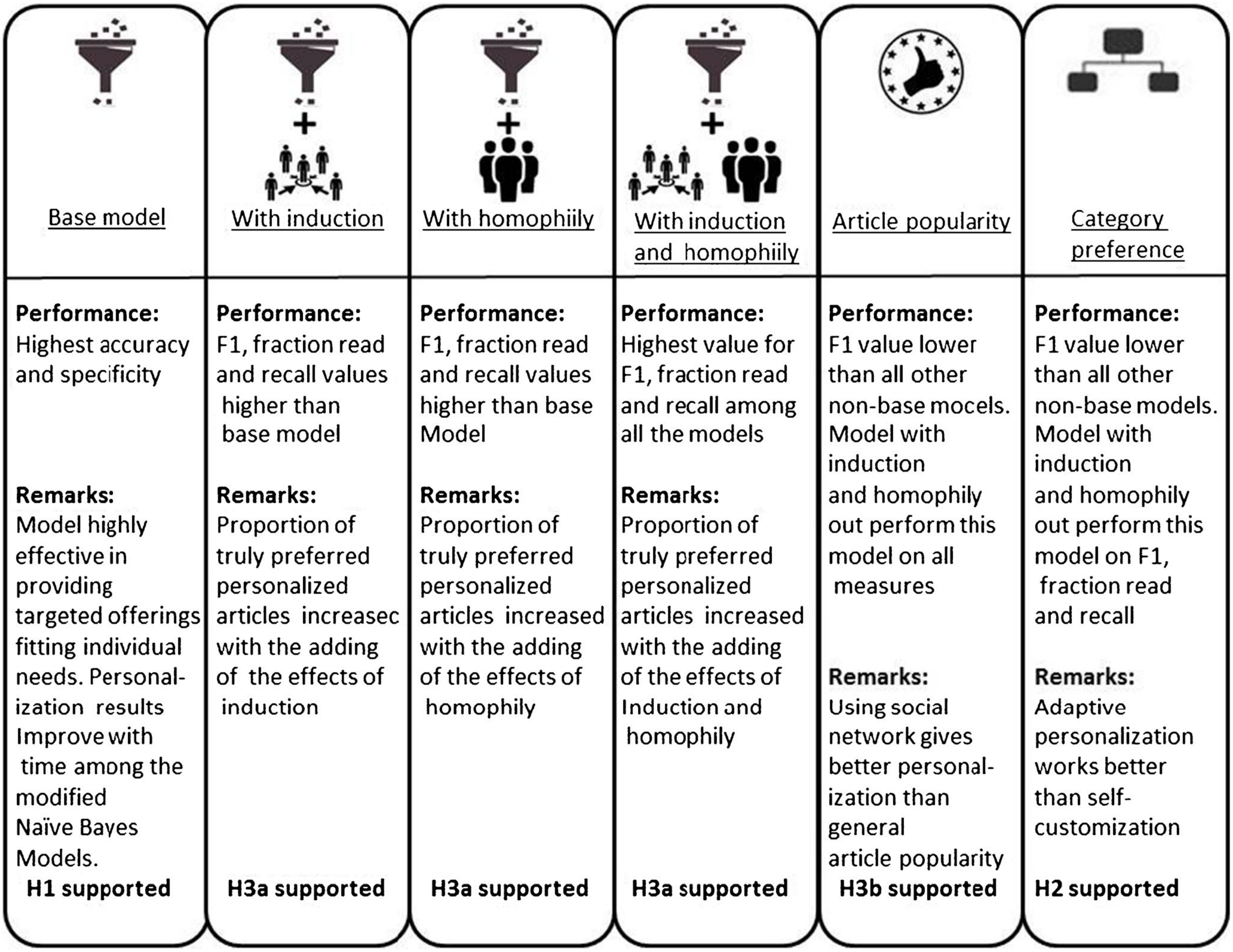


Fig. 4 Results of Study 1

personalization system was programmed in the Visual Basic .NET Compact Framework on the mobile devices. Participants were issued a mobile device (PDA) to download and read the mobile news at their own time and pace, while they went about their daily life, in "normal situations." Data collection occurred in a period of 7 to 10 days, in which participants could have multiple cycles of personalization.

During the downloading process, RSS news feeds were extracted from a major business news site. The news articles were personalized on the mobile devices as a part of the downloading process, using the AP algorithm described above. The reading histories were automatically shared in the background. The execution of the personalization took a few minutes in the first day to a few hours at the end of the experiment. The duration of the execution is dependent on the size of the word table (the collection of the words from prior read and unread articles) and on the available computing power. Due to the processing time needed on the mobile devices used in our field study, participants were encouraged to run the downloading process overnight while charging the mobile device. The initial list of articles provided in the mobile device was not customized to the preferences of the participants, as no individual reading data were yet available, but was used to initialize the procedure. Other aspects of the design of the study were similar to what was described above for study 1.

Implementation details of the personalization procedure Based on the findings of a simulation[3] we used the counts of the 55 most predictive keywords from the entire text of news articles. We restricted the number of headlines to scroll on the mobile device to be between 15 and 30. When more than 30 articles had an odds ratio, $o_r(x_{b,1:W+V})$, larger than one indicating that the articles are predicted to be preferred, the scroll-list contained those 30 articles with the

[3] A technical appendix describing the simulation is available from the authors.

highest odds and no further articles were provided from the user's social network. Conversely, when fewer than 15 articles had an odds ratio greater than 1, headlines of 15 articles with the highest odds ratio were scrolled (even if their odds ratio was less than 1), and additional articles were added from the social network to increase the number of articles in the scroll list. The articles added from the user's social network were those which at least one member of her social network had read within the last 48 h, and to which a target participant was indifferent, i.e. had an odds-ratio of $1-\varepsilon<o_r(x_{b,1:W+V})>1+\varepsilon$, with a value $\varepsilon<10^{-3}$ for the social influence parameter. This value was chosen based on some experimentation with synthetic data. By further tuning the social influence parameter at the individual level during the field study the influence of the social network can be increased and personalization might be improved. We suggest the development of methods for the selection of this parameter as an area for further research.

Results Because participants were given the freedom to decide when and how frequently they wish to acquire new news articles, the number of personalization cycles differed from one participant to another. The average number of cycles was 3.4 (a minimum of 1 and a maximum of 8). All 109 users downloaded news items at the first cycle, and 72 engaged in the median number of three cycles. Note that because the downloading cycles were temporally asynchronous across users, the number of cycles may reflect their different involvement with reading news on a mobile device. On average, users interacted with the system for more than 40 min per cycle, which provides evidence of considerable interest in the news articles downloaded from the business news website, and personalized. Only 11% of the users have more than 5 cycles, and we therefore base our analysis on the first 5 cycles.

On average, 26.3% of the personalized articles were read, and it increased by an average of 2.0% per cycle, indicating that the adaptive personalization system worked better over time, supporting H1. The average reading time per article was close to one minute (56.9 s.), and it increased by an average of 1.2 s per cycle (2%), again indicating successful adaptive personalization. In this field study, the performance of the system is substantially better than in the benchmarking study above. Here, 26.3% (i.e. TP/N) of all articles were read versus 16.4%, and the F1 value was .543 versus .446 in study 1's AP(Induction and Homophily model). The reason for this superior performance is the better and better personalization of news across subsequent cycles of personalization. Removing 5.57% cases with extremely large odds higher than 999 (the predicted probability of reading is very close to one), the average odds for the personalized news articles is 7.950, while the average odds for the articles that were not personalized was only .921, a more than an eightfold increase. These results provide more evidence that the proposed system for personalization adapts better to user preferences as users interact with the system longer.

Without utilizing the social network in the personalization, the number of articles read drops to 24.4% and the F1 value drops to .519, which shows that personalization through the social network improves performance by about 7.8%, supporting H3a. Across cycles 1 to 5, in total 399 (7.9%) articles were recommended through the social network. Of these, 54.14% were read, which is substantially higher than the 42.62% of the other personalized articles that were read. Thus the improvements achieved by the use of social network in the present study are substantial, but they are smaller than those in the benchmarking study above. The main reason for this is that the social networks in this second study were small and more spontaneously assembled, so that the reading interests of peers in the social networks were less homogeneous. Also, the personalization itself produced sufficient numbers of preferred articles, and only about 8% of articles were recommended via the social network. Nevertheless, the present results are encouraging and further support the effectiveness of the use of social networks in personalizing news.

The most frequently personalized article, "Tibetan encyclopedia shows brain surgery was practiced 2900 years ago," (a scary prospect, indeed) was read in 60% of the cases in which it was shown; participants spent about 56.4 s reading it, with about .20 s per word. The most frequently read article was "Eat popcorn, cheese for healthy aging!", which was read 93% of the cases it was shown in the news scroll. Participants spent 49.6 s reading it on average, which was .23 s per word. Articles that were read by more users were also read longer, about 3.5 s for each time the article was read more. This is evidence for both initial (based on the headline) and sustained (based on the body text) interest in these personalized articles, and supports the effectiveness of adaptive personalization.

The 20 most frequent words in the articles read by the participants, eliminating the most common words in the English language[4] were: 1. Japan, 2. Government, 3. Company, 4. China, 5. Market, 6. World, 7. Country, 8. Business, 9. City, 10. Foreign, 11. Information, 12. Minister, 13. Earthquake, 14. Police, 15. Nuclear, 16. Power, 17. Food, 18. Bank, 19. Growth, 20. Court.

This list of words reflects events which triggered the reading interest among participants. For example, the words: "Japan", "Government", "Earthquake", "Nuclear" and "Power" probably resulted from continued concerns on the status of the nuclear power plant in Japan after an earthquake. That the word "food" appears in the list as well may indicate that participants were concerned with possibility of food contamination as a result of the radioactive leakage. The participants were of voting age and their interest in the local government policies is reflected in words like: "Government", "Minister" and "Growth". Finally, their interest in stories

[4] Obtained from https://www.englishclub.com/vocabulary/common-words-100.htm

relating to crime is reflected by words like "police" and "court". The list of the most frequent words extracted from the articles read by the participants give face validity to the effectiveness of adaptive personalization algorithms. It shows that reading interests are triggered by major events and provides convincing evidence of the increasing effectiveness of the personalization as users interact with the system more: one of the key issues this research set out to investigate.

Figure 5 provides an overview of the findings from Study 2. This study shows that an adaptive personalization system can be successfully implemented in the mobile news environment, in a natural field implementation, and provides additional support for H1, that the adaptive personalization system improves its performance over time, and H3a, that utilizing the social network improves the performance of the system.

Post field studies robustness check

In the application of the system we needed to specify certain parameters of the algorithm, in particular the number of keywords to be included (W), the time window of interest (M) and the value of the social influence parameter (ε). The values of these parameters used in study 1 and 2 were selected based on a simulation study in which we simulated individual reading preferences using an individual-level logistic regression. While the optimal values of these parameters are likely context specific, the data collected in our two field studies allow us to assess the stability of the performance of the system with respect to the parameter values used. We therefore conducted another simulation study using the keywords obtained from study 1 to simulate the real preferences of 10 individuals. The values of W used were 30, 55, 80 and 110 keywords, the values of M were 4, 7, 10 and 14, and the values of ε were 0, 10^{-1}, 10^{-3} and 10^{-6}. The settings of these parameters were varied by changing the settings in the base-case one at a time (W=55, M=7, $\varepsilon=10^{-3}$). A total of 868 news articles used in study 1 were used for the simulation, involving over 8000 data points.

The result of this simulation is shown in Table 4. We use a model that has W=55, M=7 and $\varepsilon=10^{-3}$ as the benchmark base model. Note that the purpose of the simulation is not to identify the optimal values for W, M and ε, but to assess the robustness of findings for settings of these three parameters around the settings used in study 1 and 2. Table 4 shows that performance of the personalization improves with an increasing

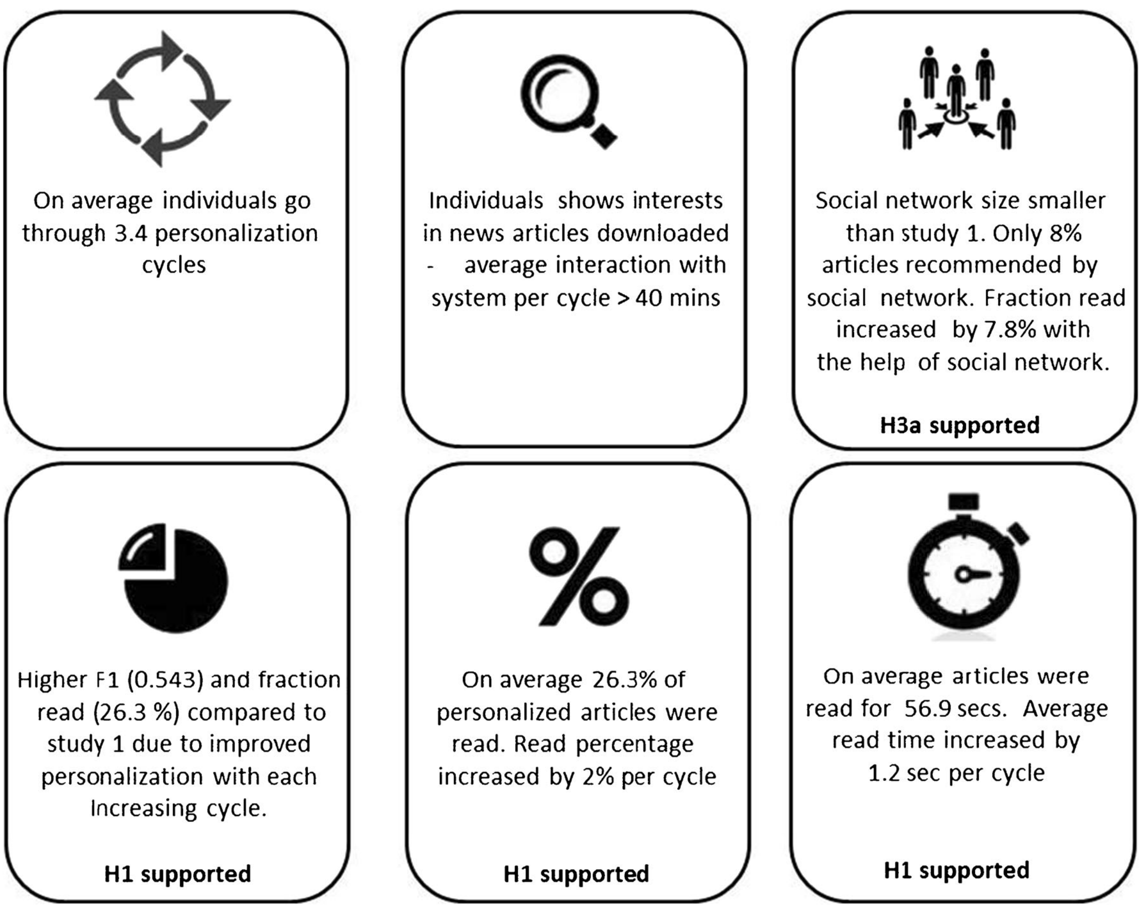


Fig. 5 Results of Study 2

Table 4 Results of the simulation for assessing robustness

No.	Method	Positive accuracy	Precision	Recall	F_1	Accuracy	Specificity	True positives (TP)	False positives (FP)	True negatives (TN)	False negatives (FN)	Total articles (N)
1	Base (W=55, M=7, $\varepsilon=10^{-3}$)	.111	.266	.998	.420	.693	.655	898	2475	4695	2	8070
2	Base with W=30	.056	.307	.500	.380	.818	.858	865	2661	4509	35	8070
3	Base with W=80	.112	.268	1.000	.423	.695	.657	900	2458	4712	0	8070
4	Base with W=110	.112	.269	1.000	.424	.696	.658	900	2450	4720	0	8070
5	Base with M=4	.111	.265	.998	.419	.692	.653	898	2485	4685	2	8070
6	Base with M=10	.111	.267	.998	.421	.694	.656	898	2469	4701	2	8070
7	Base with M=14	.111	.267	.998	.422	.695	.657	898	2462	4708	2	8070
8	Base with ε=0	.041	.433	.369	.399	.876	.939	332	434	6736	568	8070
9	Base with $\varepsilon=10^{-1}$	.111	.265	.999	.419	.691	.653	899	2491	4679	1	8070
10	Base with $\varepsilon=10^{-6}$	.111	.268	.993	.422	.697	.659	894	2443	4727	6	8070

number of keywords (W) and longer time window (M). The simulation results do demonstrate that personalization improves when ε is 10^{-3} rather than when ε is 10^{-1}. A higher value of ε moves the personalization too much away from the individual's predicted preferences. The results are quite robust to settings of ε around the value in the applications, however. All in all, these results provide evidence that the recommendations are robust to variations in the parameter settings around the ones chosen in the applications. Even though results improve with higher value of keywords (W) and longer time window (M) our simulation results indicate that the improvements from larger W and M values may not justify the resulting increases in computation time during actual field studies.

Finally, we also look at the performance of the system across cycles, and we compare day 1–7 versus the performance of day 8–15. Note that articles in day 8–15 are outside the time window of the articles that the system started with, as M=7. We use a model that is similar to the one used in the field study (i.e. a model with W=55, M=7 and $\varepsilon=10^{-3}$). Personalization performance is superior on all measures in days 8–15 (F1=.434, Recall=1.000, Precision=.277, Positive accuracy=.119) compared to the performance in day 1–7 (F1=.408, Recall=.996, Precision=.257, Positive accuracy=.105). This shows that stem performance improves with increasing cycles. Personalization performance should improve even further when more individuals go through more recommendation cycles.

Discussion

Findings and implications

This research explored the performance of an adaptive personalization system in building a better personalized product over time. To address issues surrounding the performance of adaptive personalization, an application in the context of mobile news was implemented, and a system was built and deployed such that people could use it in a natural way in their daily life.

Many criteria can be used to assess the performance of a recommendation system or personalization system. They include correctness metrics, serendipity, risk taking, reliability, response speed, and robustness to attack (Herlocker and Konstan, J.A.& Riedl, J 2000). Correctness metrics such as accuracy, precision and recall are typically used to evaluate how well a system performs, but they are insufficient for evaluating user satisfaction (McNee et al. 2006). In the context of news articles, users of the system may also be interested in serependity in order to be pleasantly surprised rather than always receiving obvious or unsurprising recommendations (Herlocker et al. 2004). The model in this paper utilizes the social network to introduce serependity in the system and to help the user and his/her network discover new articles that may be of interest (Hosanagar et al. 2013).

The learning obtained from the social network is valuable, as research shows that a user learns more about a product category from peer reviews than from the user's own usage experience (Zhao et al. 2013. The context in which the system is implemented in this paper allows the system to trade off some reliability (exploitation) for new and potentially surprising content (exploration). By contrast, a user of a recommendation system for medical and health products might be more concerned with reliability of the system's recommendations, and want to avoid random surprises.

Response speed is another factor which may affect user satisfaction and sometimes it is more important than accuracy of the recommendations. The modified Naïve Bayes model used by this system allows the system to be scalable and be implemented in real time; this improves the system response speed and by so doing may increase user satisfaction.

In addition to the criteria identified by Herlocker et al. (2000), this paper utilizes an adaptive system in which the profile of a user evolves and is updated rapidly. This is necessary in the context of news where news items and topics change quickly and where users cannot express exhaustively their interests relating to news (Picault and Ribière 2008). This explains the superior performance of the paper's model when compared against the approach of using user's predefined reading interests (self-customization), in line with the results in Franke et al. (2009).

Consideration of differences in taste matters when it comes to user preferences for products like music, news and restaurants. There needs to be a taste match between the target user and the network of peers in which the social network recommendation is applied (Yaniv 2004). Simple aggregation of preferences, as in the case of a simple popularity model, will not work as well as the model used in this paper, as it ignores differences in tastes among the different segments of users.

Several key findings emerge from this research. First, the adaptive personalization system was shown to successfully improve the product over time, supplying news articles that were read at a higher frequency (H1). The system also performed better than self-customization (H2), or simply choosing articles based on their popularity (H3b). This is an important finding, because it runs counter to most current business practice, which tends to use one of those two simpler approaches. This suggests that using an automated algorithm to personalize the product offering may be a better approach. Third, this research also shows that employing information from the customer's social network can improve personalization performance (H3a), and that this improvement could come from social influence (induction) and similarity (homophily). Although such sharing of information may seem to erode privacy to a degree (although much of the information can be shared anonymously) the practice may be a win-win, because the customers gain from better personalized products, while the service providers gain from being able to satisfy their customers better (and presumably monetize some of that added value).

Currently, social networks such as Facebook or LinkedIn allow others in a user's social network to indicate that they like or recommend a service. Most news websites, including the *New York Times* and the *Wall Street Journal*, offer a similar option, where users can 'like', or 'share' a news-article on the website with friends in their network. This, however, requires proactive effort on behalf of the user. In addition, sharing news through a social network flags the user as the source, which in some cases may not be desirable to the user and might hamper the spread of news content through the network. In our approach, the social network is employed only when the system determines that user preference is uncertain. The study shows how social networks can improve personalization even without any proactive sharing of content or observation of overt behavior of peers. But of course, our system does not preclude the proactive sharing of content or preferences as well.

The proposed system exemplifies a recent trend towards Closed Loop Marketing (CLM). CLM comprises of a cycle in which customer information is continuously and automatically collected, advanced analytics are used to predict customer behavior, and services and marketing effort are redesigned and personalized, in short cycles (Capgemini 2008; Rust and Huang 2014). Our approach involves an application of fully automated closed loop marketing to the news industry. The news industry is in transition and has experienced significant drops in revenue in the past 15 years, caused by changing consumption behavior of news content, and the rise of online and mobile news sources (Kirchoff 2010). The industry is in search of new business models, and is experimenting with various online/offline subscription schemes and revenue-sharing with online search engines. We believe that our study shows that closed loop adaptive personalization offers potential as a business model for the news industry, in particular for online and mobile news. As our field study shows, adaptive personalization will increase news consumption over time (we found an average increase of about ten percent across a mere five cycles), and thus increase the associated news revenues.

Limitations and future research

The current research includes only one application area—mobile news—and thus any generalized conclusions from the research must be considered preliminary. Likewise the implementation, although much more realistic than the typical laboratory study, had as a limitation that it did not utilize the subjects' own smart phones. It is possible that their behavior might be somewhat different in such a scenario. Another limitation has to do with the time frame used in our field studies. The system uses a Bayesian procedure based on keywords thus allowing it to improve on its estimation with bigger data sets, or to adjust its estimation with new information. We expect that the performance of the system to improve with time and to adapt to changing reading preferences but we have not tested the long term performance of our system. We recognize this as a limitation of this paper. In addition, the fundamental premise of the adaptive personalization approach is based on not requiring proactive effort on the part of the individuals. This encourages adoption of the system, as it is less obtrusive to the individual, and reduces search and cost of processing information, thereby making exploring and reading news more pleasurable and satisfactory. We have thus elected to remove the self-personalization option from our system, and from its description in the paper. There is a possibility that the system performance may improve with some automation combined with some self-personalization, however we have not explored that possibility in this paper. Future research is needed to further test the adaptive personalization approach, and devise new methods of applying it. Many new contexts appear ripe for such applications, such as other mobile services, education, entertainment and search. The psychology of how customers react to adaptive personalization is also completely unexplored.

There are many other possible avenues for future research. In the system used, assumptions are made regarding various parameters in the models (e.g., the value of the social influence parameter, the number of keywords included and time window of interest). The models in this paper stabilize fairly quickly, e.g. in the fourth round of estimation. However, it depends greatly on how stable the individual preferences are and how dynamic the news items and topics are. Possible areas of future research include (1) looking at the point of diminishing / negative return on the number of keywords used, and (2) looking at the relationship between the perceived value of the social network and individual receptiveness to reading articles outside of their usual preferences.

In addition, we would like to know how much of the customer's social network should be used. Should it be the entire network, just the person's closest links, or that part of the network to which the user has a specific type of association? Should those in the social network perceived to be closer to the user get more weight when personalizing news? Should those in the social network closer to the user's geographic location get more weight? The data collected through adaptive news personalization systems in principle enable these questions to be answered. Measures of social connectivity, e.g. centrality (Freeman 1977) can be incorporated into future studies. As these questions begin to be addressed, and advances in information systems and communications technology produce an increasing array of interactive services and devices, we expect that adaptive personalization systems, such as the one developed and studied here, will become increasingly effective and widespread, and will create substantial value for consumers and companies.

Conclusions

Our research provides further evidence that adaptive personalization can be a successful personalization approach, and that including information from the customer's social network can further improve personalization performance. We demonstrated how adaptive personalization can be successfully implemented in the context of mobile news. Our research suggests that adaptation over time can make a product better. Importantly, this approach performed better than the usual industry approaches of having the customer select desired product features (self-customization), or simply presenting the most popular articles. Adaptive personalization systems appear to have considerable potential, especially in the fast-growing area of information service.

Appendices

Appendix 1 - Procedure for filtering news feeds

To predict whether an article will be read or not, we develop a fully automated probabilistic modified Naïve Bayes approach to our Adaptive Personalization (AP) algorithm. As input, the approach uses keywords in the text of news articles that describe an individual's interests in news. Assume, for one user (with user-specific subscript suppressed for convenience), that every article is tagged as r=1 or r=0 depending on whether it is read or not. Assume there are $s_0=1,\ldots,S_0$ unread and $s_1=1,\ldots,S_1$ read news articles, with a set of W unique words across all articles. At any point in time, the data for the user then consists of a $S\times W$ table, $Y=\left(Y^0\ \ Y^1\right)$; with $Y^0=y^0_{1:S_0,1:W}$, $Y^1=y^1_{1:S_1,1:W}$, respectively, counts of unread and read words, respectively, where $(1{:}S_0,1{:}W)$ indexes a matrix. If we sum the table across $s_0=1,\ldots,S_0$ and $s_1=1,\ldots,S_1$ we have $y^0_{1:W}$ and $y^1_{1:W}$, $(W\times 1)$ vectors of word counts. This comprises the training dataset. The algorithm assumes a mixture-Poisson distribution for the joint distribution of the counts of the W keywords, with parameters $\lambda_{r,w}$ describing the intensity of keyword w in class r, which is of proportion π_r:

$$P(Y)=\sum_{r=0,1}\pi_r\prod_{w=1}^{W}\frac{\lambda_{r,w}^{y^r_w}e^{-\lambda_{r,w}}}{y^r_w!}. \tag{1}$$

A mixture-Poisson distribution model has a gamma distribution as the mixing distribution of the Poisson rate. Equation 1 is estimated on the word count data Y each time a new set of articles is to be downloaded. Let a batch of b=1, $\cdots$, B new articles be characterized by word counts $x_{1:B,1:W+V}$. There are V new words in that batch that are not in the training set. These new words are irrelevant in classifying the new article (because there is no prior data on them), but are relevant in updating the estimates after this new batch of articles, with the V new words, becomes part of the training data. The estimates of the parameters $\lambda_{r,w}$ are updated before every new downloading cycle.

Keyword selection and updating Given that we use a probabilistic classification, a coherent selection mechanism to obtain the W most discriminating keywords is based on the classification odds-ratio:

$$o_s(y_w)=\frac{P(y_w|r=1)(1-P(y_w|r=0))}{P(y_w|r=0)(1-P(y_w|r=1))}$$
$$=\frac{\lambda_{1,w}^{y^1_w}e^{-\lambda_{1,w}}\left(y^0_w!-\lambda_{0,w}^{y^0_w}e^{-\lambda_{0,w}}\right)}{\lambda_{0,w}^{y^0_w}e^{-\lambda_{0,w}}\left(y^1_w!-\lambda_{1,w}^{y^1_w}e^{-\lambda_{1,w}}\right)}. \tag{2}$$

This classification odds-ratio is large for words that discriminate well between the read (r=1) and unread (r=0) articles. It is one of the best performing procedures for keyword selection, and may reduce the number of keywords by a factor of 100 without a loss in performance (Sebastiani 2002). In our algorithm, every time a new batch of articles is downloaded and reading behavior is observed, words in all available articles are sorted by the magnitude of their odds-ratios in Eq. 2, using the data up to that point. The sorted list is then truncated at a pre-specified number of key words, W. The cutoff is set to W=55 words from the entire text, based on experimentation with synthetic data, which is detailed in Appendix 3.

At any time, a moving window of M batches of articles is included, and articles older than these M batches are removed. This helps to dynamically remove keywords that no longer sufficiently reflect a user's interests, and thus adapts the classification to the changing interests of the user. We suggest setting M at a value that reasonably reflects (1) how frequent user reading interests change, (2) how frequent and how drastically news content changes; (3) how frequently the users engage with the system. For our application, we judge that M=7, which results in about one week of articles if articles are downloaded in daily cycles by the user, weighs these three factors appropriately. M=7 yields satisfactory performance in our example (see below), but we have not tested the impact and effectiveness for other values of M. Different applications may require different values of M. For example in the case where the user accesses the system rarely, we may need a bigger value of M, because it will take longer to evaluate the user's preferences.

Keyword dependence Maximum likelihood inference for the parameters in (1) leads to the closed form estimators: $\widehat{\lambda}_{r,1:W} = y^r_{1:W}/S_r$ and $\widehat{\pi}_r = S_0/(S_0 + S_1)$. However, zero values in the case where a keyword does not occur are common. An empirical Bayes procedure derived from an indifference Gamma prior with parameters c_1 and c_2 is used to address this problem. This prior reduces the classification error, especially when little data are available (Cerquides and Màntaras 2003; Eyheramendy et al. 2003). A limitation of (1) is that it is based on the assumption that the occurrence of words is conditionally independent given the classes r. This assumption can be alleviated by averaging across all classifiers that are obtained conditional on the presence of a certain set of keywords (Webb et al. 2005). The conditional rate of occurrence of the word w, $\lambda_{1,w|z_k}$, given that a certain keyword k is present, z_k=1, or absent, z_k=0, is estimated as:

$$\widehat{\lambda}_{r,w|z_k} = \frac{c_1 + y_{r,w|z_k}}{c_2 + S_{r|z_k} + 2}, \tag{3}$$

where for class r, $y_{r,w|z_k}$ is the number of times word w appears in articles that contain or do not contain keyword k, and $S_{r|z_k}$ is the number of news articles with or without keyword k.

The joint probability of reading an article in class r that contains the keyword, z_k, $P(r,z_k)$ is estimated as:

$$\widehat{P}(r, z_k) = \frac{S_{r|z_k} + W + 1}{S_0 + S_1 + 2 \times W + 2}. \tag{4}$$

The probability of reading or not reading a new article b is then estimated as the average across equation 4 over all values z_k=0,1 for key-words k=1,…,K:

$$\widehat{P}\left(r|x_{b,1:W+V}\right) \propto \sum_{k=1}^{K} \sum_{z_k=0,1} \left(\widehat{P}(r, z_k) \prod_{w=1}^{W} \widehat{\lambda}_{1,w|z_k}^{x^r_{w,k}} e^{-\widehat{\lambda}_{1,w|z_k}} \right), \tag{5}$$

where the expression 5 is normalized by its sum over r=0 and r=1. Here, $x^r_{w,k}$ is the frequency with which word w occurs jointly with keyword k in article of class r. Substituting equations 3 and 4 yields the expression, for a specific user, of the probability of reading the new news article given the keyword count-vector:

$$\widehat{P}\left(r|x_{b,1:W+V}\right) \propto \sum_{k=1}^{K} \sum_{z_k=0,1} \left[\frac{S_{r|z_k} + W + 1}{S_0 + S_1 + 2 \times W + 2} \prod_{w=1}^{W} \left(\frac{y_{r,w|z_k} + c_1}{S_{r|z_k} + c_2 + 2} \right)^{x^r_{w,k}} \exp\left(-\frac{y_{r,w|z_k} + c_1}{S_{r|z_k} + c_2 + 2} \right) \right]. \tag{6}$$

A crisp classification of the new article is obtained as $\text{argmax}_r \widehat{P}\left(r|x_{b,1:W+V}\right)$. This classification is the basis for the filtering of news articles in the adaptive personalization system.

Including the social network As a key feature of our approach, we include the social network of a target user in the personalization. In the user's news scroll, we automatically include articles read by the user's friends that the user would otherwise not have been presented. Peer influences are stronger if an individual is less certain about his/her preferences (Narayan et al. 2011). We thus focus on articles for which the user's preference is ambivalent, because when the uncertainty of liking or disliking these articles is the highest, the potential for peer influence the largest. We thus restrict the use of the social network to only the news articles towards which a user is predicted to be indifferent; articles that the user would be predicted to read are downloaded in the mobile device anyway. This circumvents the problem of downloading articles from the social network which respondents are likely to dislike.

Thus, any news article b for which a target user is predicted to be more or less indifferent is downloaded to that user's mobile device if anyone in her social network has read it. In our model a user is predicted to be indifferent to an article when the odds of reading the news article,

$$o_r\left(x_{b,1:W+V}\right) = \frac{\widehat{P}\left(r=1|x_{b,1:W+V}\right)}{\widehat{P}\left(r=0|x_{b,1:W+V}\right)}, \tag{7}$$

is close to 1. That is: $1-\varepsilon<o_r(x_{b,1:W+V})>1+\varepsilon$, for some small value ε which ensures that the algorithm is very selective. We should emphasize that when a user is *predicted* to be indifferent to an article and the odds of reading is close to 1, the individual in question may or may not in fact be indifferent to the article. When the odds of reading an article is close to one, this may reflect uncertainty of the system about the users preference, rather than the user truly being indifferent. The parameter ε can be interpreted as a "social influence parameter", because its value controls how many of the news articles enter the users' news scroll through the social network. To be selective, the default value of the social influence parameter is set to be a small value.

This feature is expected to improve personalization if users in the social network have a shared preference for the news articles. In addition, it offers the important advantage of preventing the algorithm from zooming in on a too narrow set of news articles early on, by introducing a certain level of 'surprise'. That is, new articles are introduced in an informed manner based on the social network, which enables the algorithm to learn the user's preferences for new topics to which they would not otherwise be exposed. Without the user's ability to see the reading behavior of their peers, the influence of the social network in our system mostly operates through a mechanism of collaborative filtering. This helps in improving preference estimation, especially when the reading preferences of the peers and the target individual match well. Effects through homophily of peers in the network do occur, however, since if news articles are preferred by the peers this also increases the chance that the target user is exposed to the same article.

Appendix 2 – Randomization test

The randomization test (Lunneborg 2000) is used to test for the significance of differences in the performance of adaptive personalization models, because measures such as precision and F1, are nonlinear and have too complex a distributional form to permit a traditional test of significance. Randomization tests differ from a parametric test of significance in many respects, among them is the lack of assumption of normality or homoscedasticity. In addition, we do not calculate the test statistic for a randomization test to a tabled distribution (e.g. normal distribution) but instead we compare the results that we obtain when we repeatedly randomize the data across the comparison groups. An important assumption behind a randomization test is that the observations are exchangeable under the null hypothesis. Under that assumption, randomization tests are exact tests. The primary purpose is to get p-values, the probability of obtaining a more extreme result under the null hypothesis of no difference. Re-sampling is done without replacement. Here the purpose is to get p-values for differences in the estimated values of F1, and the other measures. between the different personalization models. Because respondents were randomly assigned to conditions, the assumption of exchangeability likely holds. Comparing the difference between bootstrapping and randomization test, the bootstrap-based is based on less strict assumptions, but it is not an exact test. It is primarily used to obtain confidence intervals, and is based on sampling with replacement Good (2005).

The null hypothesis is that two models are not different, and that therefore any prediction produced by one model could have just as likely come from the other model. In the randomization test, we shuffle the predictions between two models and see how likely this produces a difference in the performance measures that is at least as large as the difference observed from the data. To illustrate how the test works, take the example where model 1 has a better performance than model 2. Then, we would expect that the performance metric obtained by randomly shuffling the predictions of the two models would not to give a larger difference in performance. In other words, a mixture of the predictions from model 1 and 2 is unlikely to give better predictions than the predictions coming from model 1 alone. In using 10, 000 randomizations, we reshuffle the predictions of model 1 and 2 10,000 times. The test produces the number of times that there is a larger difference in performance as compared to the observed data for models 1 and 2. The test results provide the proportion, or probability, that the randomized predictions are better than the observed predictions. This probability can be interpreted as a 'p-value' measuring the level of significance of differences between the models.

The steps involved in a randomization test are:

1. Select on a metric for comparison between the different personalization models (e.g. F1)
2. Calculate the metric based on the different models. We denote the personalization model with a better performance as system 1 and the other one as system 2.
3. Repeat the following 10,000 times (a value we chose for this study):

    a. Shuffle the data between the two models, that is shuffle the predictions for the different individual news articles between the models,

   b. Calculate the metric for the shuffled data,
   c. With the shuffled data, if the performance of model 2 is better than model 1 increase the value of a counter by 1,
   d. Continue step a to c 10,000 times.
4. Divide the value of the increment counter by 10,000 to get the proportion of times the performance of the model 2 is better than model 1 with the shuffled data.

   Reject or retain the null hypothesis on the basis of this probability.

Appendix 3–tests on simulated data

We conduct a simulation study to investigate the performance of the Naive Bayes algorithm for various amounts of data used as input. For this purpose, a total of 482 news articles which were downloaded from Channel News Asia Business RSS feeds between 1st February 2010 and 9th February 2010, are used for the simulation. These articles span a period of 7 days. The moving window from which articles are retained is set to $M=7$, since most articles lose much of their news value after 7 days. The news categories of these articles are Asia Pacific (69 articles), Business (69), Entertainment (70), Local (70), Sports (70), Technology (66) and World (68). A discriminant analysis is used to see which the keywords helped to discriminate the news articles. The recommended ratio of observations to predictor variables is at least 5:1 (Hair et al. 2005), so that with 482 observations we select the 96 keywords with the highest frequency.

We use the counts of these keywords in an individual-level logistic regression for ten individuals, to simulate their reading preferences. The coefficients are drawn from a uniform distribution. The logistic regression produces an indicator variable that shows each of the 482 articles to be either 'read'(0) or 'skipped' (1) for each individual, and this is matched against the prediction of our algorithm. The articles from the first day are used to initialize the procedure (i.e. to generate the keywords and the word tables for read and skipped articles). The keywords and word tables generated from the first day are used to personalize the articles for day two, and so on. A total of 412 articles is available. The social network is simulated by assuming that peers in the social network have identical reading preferences, up to a random error. Because we are interested in personalizing the articles read on an individual) basis, we simulated 10 users.

The classification performance of the approach depends critically on what and how much data is used as input. As there are no theoretical guidelines available for these choices, we test the performance of our approach using four alternative approaches:

1. The words in the headlines of the news articles,
2. The words in the first three paragraphs of the news articles,
3. The words in the entire contents of the news articles,
4. The entire contents of the articles and incorporating the social network.

Performance will depend on how many keywords are used. For each of the scenarios, we therefore compare the performance of the approach as a function of the maximum number of keywords used for classification, varying it from 25 to 55 in increments of 10. The maximum of 55 keywords is dictated by memory and processing limitations of the mobile devices. The values of the F1 measure for the resulting scenario's are shown in Table 5.

The motivation for testing our algorithm using only the headlines as data comes from the observation that users only get to see the headlines before deciding if they want to read the news articles. Due to the small number of words in the headline, F1 performance does not improve by increasing the maximum number of keywords. In fact when from 25 to 55 keywords are used no articles are recommended. This results in a zero value for precision and recall which resulted in a zero value for F1. In addition, with no articles recommended (i.e. no positives), the measure of accuracy simplifies to a simple ratio of true negatives divided by total negatives. This together with the simulated users preferring 11% of the articles on average, resulted in an artificially high accuracy. Obviously, this performance is unacceptable and we conclude that only using the headlines does not yield effective personalization.

The motivation for testing the performance of the algorithm using only the keywords in the first three paragraphs of each news article, is that users have a tendency to browse through articles and classify a document as interesting after reading only the first few paragraphs, without scrolling. Using 25 keywords in this case yields F1=.207. This is some improvement over using keywords from the headlines only. Yet, when the number of keywords is increased, performance decreases, ultimately to F1=.019 for 55 keywords. Inspection of the data shows that this is due to the algorithm's tendency to reduce the number of articles recommended as the number of keywords increases. For this reason, we conclude that using the first three paragraphs does not yield acceptable performance either.

When keywords from the entire contents of the news articles are used, the performance of the algorithm improves with the number of keywords. Using from 25 to 55 keywords F1 improves from .206 to .255. Inspection of the data shows that this is caused by the algorithm being good at screening out

Table 5 F1 Performance measures for the simulated data

# Keywords:	25	35	45	55
Headline	.000	.000	.000	.000
First three paragraphs	.207	.101	.059	.019
Entire text	.206	.223	.257	.255
Entire text plus social network	.317	.416	.570	.619

unwanted news articles. But the algorithm is less good at finding news items that are preferred by the user. Compared to the previous two scenarios, however, these results are the best. These simulations show that there is little opportunity to improve computational performance by reducing the data input: for satisfactory performance one needs the entire news article as data, using the maximum number of 55 keywords to classify the article.

Finally, when not only the entire content of the news articles, but also the behavior of friends in the social network of the target user is used, F1 improves monotonically from .317 (25 keywords) to .619 (55 keywords) (Table 5), which is substantially better than with the same number of keywords without the social network. Breaking down performance by recommendation cycle shows that the F1 value increases from .614 for the first cycle, to .662 to the sixth and last cycle, revealing that the approach adapts better to user interests over time. Based on these results we decide on using 55 keywords from the entire text, as well as the social network, in the algorithm.

References

Adomavicius, G., & Tuzhilin, A. (2005). Personalization technologies: A process-oriented perspective. *Communications of the ACM, 48*(10), 83–90.

Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research, 37*(3), 363–375.

Ansari, A., & Mela, C. F. (2003). E-customization. *Journal of Marketing Research, 40*(2), 131–145.

Ansari, A., Koenigsberg, O., & Stahl, F. (2011). Modeling multiple relationships in social networks. *Journal of Marketing Research, 48*(4), 713–728.

Atahan, P., & Sarkar, S. (2011). Accelerated learning of user profiles. *Management Science, 57*(2), 215–239.

Bechwati, N. N., & Xia, L. (2003). Do computers sweat? The impact of perceived effort of online decision aids on consumers' satisfaction with the decision process. *Journal of Consumer Psychology, 13*(1&2), 139–148.

Bell, D. R., & Song, S. Y. (2007). Neighborhood effects and trial on the Internet: Evidence from online grocery retailing. *Quantitative Marketing and Economics, 5*(4), 361–400.

Capgemini. (2008). *Closed loop marketing: Unlocking the benefits of customer centricity*. http://www.us.capgemini.com/lifesciences/. Accessed 18 Aug 2011

Caumont, A. (2013). *12 trends shaping digital news*. http://www.pewresearch.org/fact-tank/2013/10/16/12-trends-shaping-digital-news/. Accessed 1 Nov 2013.

Cerquides, J. & Màntaras, R.L. (2003). The indifferent naïve bayes classifier. *Proceedings of the 16th International FLAIRS Conference.*

Chung, T. S., Rust, R. T., & Wedel, M. (2009). My mobile music: An adaptive personalization system for digital audio players. *Marketing Science, 28*(1), 52–68.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). New York: Harper Collins.

Cohen, W. W., & Singer, Y. (1999). Context sensitive learning methods for text categorization. *ACM Transactions on Information Systems, 17*(2), 141–173.

Cooke, A. D. J., Sujan, H., Sujan, M., & Weitz, B. A. (2002). Marketing the unfamiliar: The role of context and item-specific information in electronic agent recommendations. *Journal of Marketing Research, 39*(4), 488–497.

Das, A., Datar, M., Garg, A., & Rajaram, S. (2007). Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 271–280).

Edmonds, R., Guskin, E., Mitchell, A. & Jurkowitz,M. (2013). *The state of the news media 2013*. http://stateofthemedia.org/2013/newspapers-stabilizing-but-still-threatened/newspapers-by-the-numbers/. Accessed 1 Nov 2013.

Eyheramendy, S., Lewis, D. D., & Madigan, D. (2003). On the naïve bayes model for text categorization. *Proceedings of the 9th International Conference on Artificial Intelligence and Statistics* (AISTATS).

Fichman, R. G., & Cronin, M. J. (2003). Information-rich commerce at a crossroad: Business and technology adoption requirements. *Communications of the ACM, 46*(9), 96–102.

Franke, N., Keinz, P., & Steger, C. J. (2009). Testing the value of customization: When do customers really prefer products tailored to their preferences? *Journal of Marketing, 73*(5), 103–121.

Freeman, L. C. (1977). Set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41.

Godes, D. (2011). Invited comment on 'Opinion leadership and social contagion in new product diffusion'. *Marketing Science, 30*(2), 224–229.

Godes, D., & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field study. *Marketing Science, 28*(4), 721–739.

Good, P. (2005). *Introduction to statistics through resampling methods and R/S-PLUS*. Hoboken: Wiley.

Google Play (2014). https://play.google.com/store/apps/details?id=com.news360.news360app&hl=en. Accessed 19 Aug 2014

Hair, J. F., Black, B., Babin, B., & Anderson, R. (2005). *Multivariate data analysis* (6th ed.). Saddle River: Prentice Hall.

Hand, D. J., & Yu, K. (2001). Idiot's Bayes: Not so stupid after all? *International Statistical Review, 69*(3), 385–398.

Hauser, J. R., Liberali, G., & Urban, G. L. (2014). Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Management Science, 60*(6), 1594–1616.

Hauser, J. R., Urban, G. L., Liberali, G., & Bruan, M. (2009). Website morphing. *Marketing Science, 28*(2), 202–223.

Hartmann, W. R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science, 29*(4), 585–601.

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

Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environment: The effects of interactive decision aids. *Marketing Science, 19*(1), 4–21.

Herlocker, J.L., Konstan, J.A.& Riedl, J. (2000). Explaining collaborative filtering recommendations. *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, 241–250.

Herlocker, J. L., Terveen, L. G., Konstan, J. A., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on information systems, 22*, 5–53.

Hosanagar, K., Fleder, D., Lee, D., & Buja, A. (2013). Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science., 60*(4), 805–823.

Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science, 30*(2), 195–212.

Karr, D. (2014). The State of Mobile in the US. https://www.marketingtechblog.com/mobile-usa-2014/. Accessed 24 Oct 2014.

Keath, J. (2011). *Instagram becomes the largest mobile social network.* http://socialfresh.com/instagram-largest-mobile-social-network/. Accessed 17 Dec 2011.

Khan, R., Lewis, M., & Singh, V. (2009). Dynamic customer management and the value of one-to-one marketing. *Marketing Science, 28*(6), 1063–1079.

Kirchoff, S. M. (2010). *The U.S. Newspaper Industry in Transition*. Congressional Research Service Report for Congress, 7–5700, www.crs.gov.

Levin, D., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science, 40*(11), 1477–1490.

Li, L., Wang, D. D., Zhu, S. Z., & Li, T. (2011). Personalized news recommendation: A review and an experimental investigation. *Journal of Computer Science and Technology, 26*(5), 754–766.

Liang, T. P., Yang, Y. F., Chen, D. N., & Ku, Y. C. (2008). A semantic-expansion approach to personalized knowledge recommendation. *Decision Support Systems, 45*(3), 401–412.

Liu, D. R., Tsai, P. Y., & Chiu, P. H. (2011). Personalized recommendation of popular blog articles for mobile applications. *Information Sciences, 181*(9), 1552–1572.

Liu, J. H., Dolan, P. & Pedersen, E. R. (2010). *Personalized news recommendation based on click behavior.* http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/35599.pdf.. Accessed 13 Aug 2014.

Lunneborg, C. E. (2000). *Data analysis by resampling: Concepts and applications*. Belmont: Duxbury Press.

Lyytinen, K., & Yoo, Y. J. (2002). Research commentary: The next wave of nomadic computing. *Information Systems Research, 13*(4), 377–388.

Lyytinen, K., Yoo, Y. J., Varshney, U., Ackerman, M., Davis, G., Avital, M., Robey, D., Sawyer, S., & Sorensen, C. (2004). Surfing the next wave: Design and implementation challenges of ubiquitous computing. *Communications of the Association for Information Systems, 13*(40), 697–716.

McNee, S.M., Riedl, J., Konstan, J.A. (2006). Accurate is not always good: How accuracy metrics have hurt recommender systems. *CHI'06 extended abstracts on Human factors in computing systems*, 1097–1101.

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

Moon, S., Gary, J., & Russell, G. J. (2008). Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science, 54*(1), 71–82.

Nair, H., Manchanda, P., & Bhatia, T. (2010). Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research, 47*(5), 883–895.

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

Newspaper Association Of America. (2014). Business model evolving, circulation revenue rising. http://www.naa.org/Trends-and-Numbers/Newspaper-Revenue/Newspaper-Media-Industry-Revenue-Profile-2013.aspx. Accessed 6 Nov 2014

Picault, J., Ribière, M. (2008). Method of adapting a user profile including user preferences and communication device. *European Patent EP08290033*.

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association, 96*(453), 20–31.

Rust, R. T., & Huang, M. H. (2014). The service revolution and the transformation of marketing science. *Marketing Science, 33*(2), 206–221.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys, 34*, 1–47.

Shapira, B., Shoval, P., Meyer, J., Tractinsky, N., & Mimran, D. (2009). ePaper: A personalized mobile newspaper. *Journal of the American Society for Information Science and Technology, 60*(11), 2333–2346.

Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research, 7*(2), 99–111.

Smith, D., Menon, S., & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing, 19*(3), 15–37.

Subramanian, C. (2012). *Gadget users hungrier for news, but media still lags in profit*. http://techland.time.com/2012/03/20/124596/. Accessed 21 Mar 2012

TechCrunch (2014). http://techcrunch.com/2013/08/13/now-with-7-million-installs-dripplers-personalized-mobile-news-app-gets-an-upgrade/. Accessed 19 Aug 2014.

Thompson, D. V., Hamilton, R. W., & Rust, R. T. (2005). Feature fatigue: When product capabilities become too much of a good thing. *Journal of Marketing Research, 42*(4), 431–442.

Urban, G. L., Liberali, G., MacDonald, E., Bordley, R., & Hauser, J. R. (2014). Morphing banner advertising. *Marketing Science, 33*(1), 27–46.

Van Rijsbergen, C. J. (1979). *Information Retrieval*. London: Butterworth.

Van Roy, B., & Yan, X. (2010). Manipulation robustness of collaborative filtering. *Management Science, 56*(11), 1911–1929.

Varki, S., & Rust, R. T. (1998). Technology and optimal segment size. *Marketing Letters, 9*(2), 147–167.

Wang, J., Aribarg, A., & Atchadé, Y. F. (2013). Modeling Choice Interdependence in a Social Network. *Marketing Science, 32*(6), 977–997.

Webb, G. I., Buughton, J. R., & Wang, Z. H. (2005). Not so naïve bayes: Aggregating one-dependence estimators. *Machine Learning, 58*(1), 5–24.

Yaniv, I. (2004). Receiving other peoples' advice: Influence and Benefit. *Organizational Behavior and Human Decision Processes, 93*(1), 1–13.

Ying, Y. P., Feinberg, F., & Wedel, M. (2006). Leveraging missing ratings to improve online recommendation systems. *Journal of Marketing Research, 43*(3), 355–365.

Zhang, J., & Krishnamurthi, L. (2004). Customizing promotions in online stores. *Marketing Science, 23*(4), 561–578.

Zhang, J. J. (2010). The sound of silence: Observational learning in the U.S. kidney market. *Marketing Science, 29*(2), 315–335.

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science, 32*(1), 153–169.