The Differences between Recommender Technologies in their Impact on Sales Diversity

Completed Research Paper

Christian Matt
LMU Munich
Ludwigstr. 28,
80539 Munich, Germany
matt@bwl.lmu.de

Thomas Hess
LMU Munich
Ludwigstr. 28,
80539 Munich, Germany
thess@bwl.lmu.de

Christian Weiß
LMU Munich
Ludwigstr. 28,
80539 Munich, Germany
weiss.christian@campus.lmu.de

Abstract

Recommender systems are frequently used as part of online shops to help consumers browse through large product offerings by recommending those products which are the most relevant for them. Although consumers' interactions with recommender systems have been subject to substantial research, it is still unclear what the effect on aggregated sales diversity is, i.e. whether this leads to predominance of fast-selling or niche products. It is also unclear, whether any potential effects would differ between specific recommender technologies.

We created a realistic web-experiment to monitor consumer behavior while purchasing digital music tracks when different recommender technologies are present. To analyze potential changes in sales diversity we used the Gini coefficient as well as additional measures. We found that sales diversity increases for all recommender technologies, except for bestseller lists. Furthermore, the differences across recommender technologies are rather small. Our findings have significant implications for online retailers and for suppliers.

Keywords: Recommender Systems, Sales Diversity, Gini Coefficient, E-Business, Digital Music

Introduction

Today, there are many online shops that easily surpass any physical retail stores in the number of different products offered. Within these online shops consumers can find virtually any product they may be looking for. The invention of products that can be fully digitized (e.g., music or books) has further boosted the tendency towards stores with a larger variety of products, since very little variable costs for stocking digital products incur. Therefore, even if most products may be sold infrequently, the aggregated revenue for this "long tail" of niche products can still maintain a profitable business (Anderson 2009). Among others, a famous example is the Apple iTunes Store with more than 26 million songs and over 700.000 apps (Apple 2013).

For consumers, the wide range of products on offer can also bring its own disadvantages since it can lead to information overflow. Therefore, nowadays, most online shops employ recommender systems (recommenders) to help their customers when browsing through their products (Adomavicius and Tuzhilin 2005). Recommenders should simplify purchase processes and suggest to consumers the products that best suit their tastes (Schafer et al. 1999).

Online retailers profit from recommenders as they encourage consumers to purchase more. However, although it has been shown that recommender systems can increase sales (Hinz and Eckert 2010), it is still not clear what their impact on sales diversity is, which is an aggregate measure that reflects the concentration of market shares across an online shop's product assortment (Fleder and Hosanagar 2009). Even in times of digital products, this question may be of great importance for online retailers and for suppliers of products. For online retailers, higher volume discounts from certain suppliers and easier efforts to concentrate marketing activities on a smaller number of products may be potential reasons. For suppliers of products a stronger focus on certain items may help to save costs. On the other hand, a broad range of products and effective tools to help make customers aware of the large product diversity may imply a competitive advantage against other online competitors. In addition, this may also imply a competitive advantage against physical retailers, for which keeping a larger range of stock is much more cost intensive.

Previous research has provided conflicting evidence on whether recommender systems foster blockbuster or long tail markets. One group of researchers believes that recommender systems help consumers to find products that they would otherwise not have found and thus increase the sales of niche products (e.g., Brynjolfsson et al. 2006). The other group holds that recommenders make already popular products even more popular and thus decrease sales diversity (Fleder and Hosanagar 2009; Mooney and Roy 2000).

To further analyze this question we implemented a very realistic web-experiment in which we created an online shop for digital music tracks. We believe that digital media products are of particular interest, since media markets have traditionally been blockbuster markets and may, due to new technologies, now shift to long-tail-oriented markets (Brynjolfsson et al. 2011). Furthermore, due to their being experience goods, media products are supposed to profit more from recommender systems, since their quality is difficult to evaluate prior to their consumption (Senecal and Nantel 2004).

With our research we aim to answer the following two questions:

1) Do recommender technologies have an impact on sales diversity?

2) If so, does the potential impact vary across different recommender technologies?

For this, we implemented several recommender technologies and measured potential changes in sales diversity by using the Gini coefficient as well as additional indicators to confirm the results. Also because of the challenges to study the impact of recommender in a real-word setting (Kumar and Benbasat 2006), we believe that an experiment is suitable to answer these research questions.

The rest of the paper is organized as follows: In the next section we present a review of the current body of knowledge in this area. Next, we introduce our research model and the measures we use. We then develop our hypotheses and describe the concrete implementation as a web-experiment. The results of our study are presented and discussed thereafter. Finally, we outline further implications and a summary of the results and conclude with the limitations of the paper.

Prior Work

Recommender systems use algorithms to filter a selection of products out of a larger number of products in order to find those that are supposed to be the most relevant for each user. The main underlying technologies for recommender systems are: Content-based filtering (CBF) and collaborative filtering (CLF) (Burke 2000; Xiao and Benbasat 2007). However, there are various sub-forms of CBF- and CLF-recommenders and also quite frequently, a hybrid combination of both methods is used to combine the strengths of both approaches (Burke 2002).

Content-based filtering compares products based on their content or their characteristics, such as artist or genre for instance (Basu et al. 1998). It aims to find products that best match the user's preferences, while these can be based on previously purchased or recommended products or also based on explicitly stated preferences (Balabanović and Shoham 1997). One potential drawback in the context of media products is that their product characteristics are difficult to describe and to classify automatically. In addition, CBF is less suitable for cross-selling, i.e. making recommendations across product types.

Collaborative filtering neglects the characteristics of the products in question, but instead recommends products based on the similarity between users and their common interests (Breese et al. 1998). CLF analyzes and compares user profiles to find the most similar users. The main idea is that if two users have stated similar preferences or purchased many products in common it is more likely that each of the users will also be interested in the products that only the other user has rated or purchased. In order to compare user profiles there are several similarity measures, such as the cosine similarity or the Pearson correlation coefficient (Wang et al. 2006). Nowadays, many websites for media products use forms of CLF-recommenders, such as Amazon.com (Linden et al. 2003) or Last.fm. Still, CLF-recommenders may suffer from the cold-start-problem, i.e. there need to be a certain amount of product and user data in order to work well (Schein et al. 2002).

A substantial number of research papers, especially in computer science, consider how to optimize recommender algorithms for scalability and more precise recommendations (e.g., Koren et al. 2009; Sarwar et al. 2000). Other work deals with the question of how to design and adopt recommender systems for business use (e.g., Ansari et al. 2000; Bodapati 2008). However, there is significantly less research on the effects of recommender systems on markets. Most of the previous studies suggest a positive effect of recommender systems on sales, for instance in the field of online bookshops (Shani et al. 2005), a mobile Internet platform (Jannach and Hegelich 2009) or an online-supermarket (Dias et al. 2008). Instead of additional sales, recommender systems can also lead to substitution effects (Hinz and Eckert 2010). In this case, the overall sales amount remains the same, just the distribution of sales across products changes. Here it is assumed that recommenders help consumers to find better matching products.

However, there is considerable disagreement and conflicting evidence within the research community on whether recommender systems lead to an increase or a decrease in sales diversity. Brynjolfsson et al. (2010) state that the key question in this field, whether recommender systems produce blockbuster or long-tail markets, is yet to be answered and see this as an excellent opportunity for further research. Brynjolfsson et al. (2011) hold that, by reducing search costs, recommender systems lead to an increased market share of niche products and thus to an increase in sales diversity. The reduction of search costs for niche products due to the usage of recommender systems is also confirmed by Hinz and Eckert (2010). They also hold that top-seller lists support the emergence of blockbuster markets. Tucker and Zhang (2011) also find empirical evidence that niche products profit from popularity lists. Oestreicher-Singer and Sundararajan (2012) have analyzed product recommendation networks and shown that they amplify the influence of complementary products by up to a factor of 3 (Oestreicher-Singer and Sundararajan 2012). Furthermore, the authors hold that an increase in the influence of these recommendation networks is associated with flatter distributions of revenue and demand.

In contrast to this, Mooney and Roy (2000) argue that recommender systems increase the popularity of already popular products and thus reduce sales diversity. In line with this, Fleder and Hosanagar (2009) show with a simulation that CLF-recommenders can have a positive influence on sales diversity on an individual level, but a negative influence on an aggregate level. Based on field data and a simulation (Fleder et al. 2012) confirm the previous assumption and again hold that recommenders lead consumers to become more similar in their purchasing habits. They further differentiate between product mix and volume effects; the first one is that recommenders cause users to have more purchases in common, while the second states that recommenders lead consumers to purchase more, thus increasing the probability that they have more sales in common.

One potential factor for the disagreement in this key question may be the variety of recommender technologies that have been used across these different papers. A first basis for our work, which explicitly distinguishes between different forms of recommender technologies, is set by Hinz et al. (2011) who found that search tools can increase or decrease sales diversity dependent on the applied search technology. In addition, what many of the papers that use field data have in common (e.g., Fleder et al. 2012; Szlavik et al. 2011) is that a large proportion of their findings relies on simulations that have been applied retrospectively on the field data. Since it is very difficult to gain access to real recommender systems and online shops and to apply the relevant changes in order to conduct research, we believe that a web-experiment provides us with an environment in which we can actively control and change parameters in order to isolate certain effects and to monitor realistic consumer behavior. In order to better illustrate potential effects and how changes may accrue, we present our research model in the following section. To the best of our knowledge, our paper is the first that uses experiments to analyze whether recommender systems increase or decrease sales diversity and thereby differentiates between different recommender technologies.

Research Model

Our research model builds upon the individual consumer's decision to purchase music tracks, while being subject to one of the five treatments: a CLF-recommender (r_{CLF}), CBF-recommender (r_{CBF}), a bestseller list (r_{BS}), a "pseudo" control group with random recommendations (r_{RD}) and a baseline treatment without a recommender (cp. Figure 1). We are aware that there are various other types of recommender systems. However, due do the experimental methodology we apply, we decided to restrict our implementation on these 5 forms of recommender systems since we assume, these are the basic forms of the most popular implementations in commercial settings.

It is important to note that since we aim to analyze changes in the aggregated measure sales diversity, we do not focus on the individual's purchase decision, as it is only a small piece of the aggregated sum of all purchase decisions that finally constitute the sales diversity. In line with this, a consumer c_i has to take the purchase decision $d_{i,l}$, in which he or she is supported by a recommender system r_j that issues the recommendation $x_{i,j,k}$. Consumers can now accept the recommendation and purchase the product $p_{i,\ r}$ which was suggested by the recommender system. However, they might also not accept the suggestion of the recommender system and purchase a different product $p_{i,\bar{r}}$. In any case, the sales diversity G_j for a market with a recommender r_j is calculated based on the cumulated market share of all products that are offered in this market, no matter whether they have been recommended to consumers or not. Therefore the influence of two recommender systems on sales diversity may also vary, even if their recommendations are the same, just because consumer acceptance of the recommender may differ (which could, for instance, be due to a more salient presentation of one of the two recommender systems or the degree of perceived personalization, cp. Komiak and Benbasat 2006). We use the Gini coefficient for our measurement and explain the underlying concept subsequently.

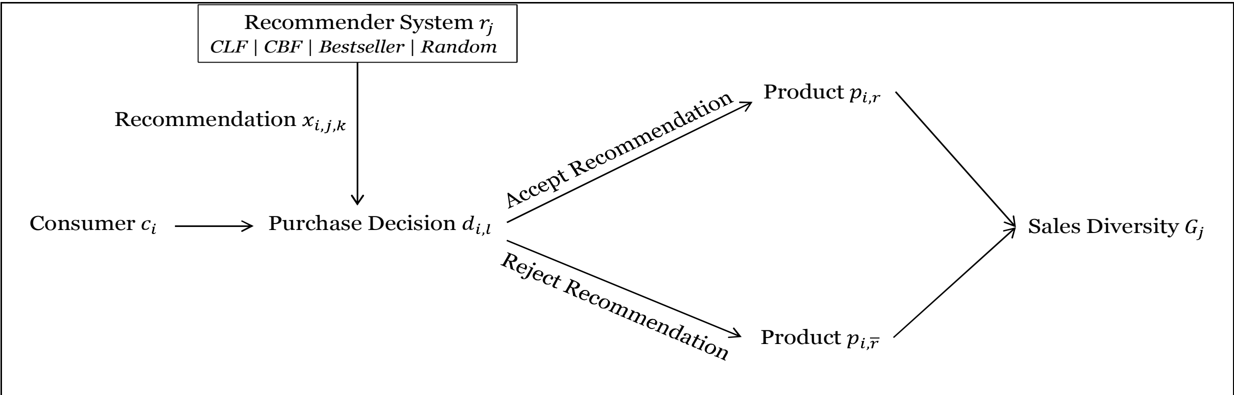


Figure 1. Underlying Research Model

The Gini Coefficient as a Measure for Sales Diversity

For our research, we define sales diversity as the statistical distribution of market shares of all products that are offered by an online vendor. A low sales diversity means that a comparatively small share of products is responsible for a large number of sales and that "niche products" (products that are sold rarely) together only account for a small number of sales. In contrast to this, high sales diversity means that sales are fairly equally distributed among all products, i.e. all products account for a similar amount of revenue.

The Gini coefficient is a common measure to quantify equalities and inequalities (Gini 1912). It is also an appropriate measure for quantifying the degree of the sales diversity (Brynjolfsson et al. 2010).

The Gini coefficient can be extracted from the Lorenz-curve as follows: Let A be the area under the bisecting line and B the area under the Lorenz curve $L(X)$, then the Gini coefficient can be calculated as follows:

$$G = \frac{A - B}{A} = \frac{0{,}5\ - B}{0{,}5} = 1 - 2 \cdot B$$

If the function $Y = L(X)$ represents the Lorenz curve, the general formula for calculating the Gini coefficient is:

$$G = 1 - 2 \cdot \int_0^1 L(X)dX$$

The Gini coefficient maps inequality distributions into a single value with a minimal bound of 0 (equal distribution across products) and a maximum bound of 1 (equals monopoly case). The transformation of the Lorenz curve to a single value is convenient for quickly assessing changes in distribution equality. However, due to this transformation, some information is hidden, i.e. in particular it may happen that different Lorenz curves have the same Gini coefficient.

To analyze the effects of recommenders on sales diversity we first calculate the Gini coefficient for a baseline treatment where no recommender system (r_0) is present. For each of the other treatments with a different recommender system $r_j$we also calculate the Gini coefficient, leaving all other factors constant. Let G_0 be the Gini coefficient for the treatment with no recommender system and G_j the Gini coefficient for the sales diversity, where the recommender system r_j is used. Then, there are three potential effects of the recommender system r_j on sales diversity G_j, which can be expressed by a comparison of the value of G_j and G_0 (Fleder and Hosanagar 2009):

Decrease in sales diversity:	$G_j > G_0$
Increase in sales diversity:	$G_j < G_0$
No change:	$G_j = G_0$

Analysis of Cumulated Market Share of Niche Products

As mentioned before, due to the characteristics of the Gini coefficient, it is possible that two different sales distributions share similar or even the same Gini coefficient. Therefore, concrete shifts of the Lorenz curve may be hidden. Therefore, it may not be sufficient just to compare the Gini coefficients, but also to have a look at the Lorenz curves. However, since Lorenz curves are difficult to compare statistically, we also extract an additional measure from the Lorenz curves that states how large the cumulated market share of the 95% slowest-selling products (the "niche products") is. An exemplary Lorenz curve is illustrated in Figure 3. The 95%-niche products measure would cover the total sales share for the area of 0-95% of the product sales share on the x-axis. We calculate this measure in addition to the Gini coefficient and use it as a support to assess whether certain recommender technologies rather lead to blockbuster or niche markets.

Development of Hypotheses

As previously mentioned, various studies generally assume that recommender systems have an effect on sales diversity, but find contradictory evidence on whether the influence of recommenders is positive or negative. This is probably due to the differences in recommender technologies as well as the methodologies these studies use. However, the assumption that recommenders have an effect on sales diversity at all is essential for any further analysis of potential differences between recommender technologies. We therefore need to test whether this fundamental assumption holds in our experimental environment which offers a high level of control. We thus pose the following Hypothesis 1:

H1: Recommender systems have an impact on sales diversity: $G_{r_{CLF}} \neq G_0$, $G_{r_{CBF}} \neq G_0$, $G_{r_{BS}} \neq G_0$, $G_{r_{RD}} \neq G_0$

Furthermore, except for Hinz et al. (2011), who analyze different search technologies, most studies do not distinguish between different recommender technologies and merely take a very basic or a given recommender technology from praxis into account. We believe that there are differences across recommender technologies and their impact on sales diversity both regarding the strength of the effect as well as the direction.

A CLF-recommender can lead to a decrease in sales diversity since popular products are recommended more often by CLF-recommenders than other products (Fleder and Hosanagar 2009). This is due to a self-enhancing circle that can arise, because products that are purchased by many customers ("popular products") are more likely to be recommended by CLF-recommenders. By assuming that the probability for a recommended product to be purchased is higher than for a non-recommended product, this may lead to the previously mentioned circle that already popular products are recommended more often and thus purchased more often and thus become even more popular. This can be seen as the main driver for a decrease in sales diversity by using CLF and which leads us to Hypothesis 2a:

H2a: CLF-recommenders lead to a decrease in sales diversity: $G_{r_{CLF}} > G_0$

CBF-recommenders promote products based on their product characteristics and not based on their popularity and the similarities between users. Therefore, products that may not be purchased very frequently may still be recommended if they have the closest fit to the consumer's product preferences. Therefore, CBF-recommenders could lead consumers to new products which they would probably otherwise not find and thus increase sales diversity. However, CBF-recommenders may be less suitable for media products as these, as being experience goods, are more difficult to describe than search goods. If this holds, the quality of the recommendations is likely to be worse than for other recommenders, which is why consumers may rather follow their natural purchase patterns and ignore recommendations (thus sales diversity would remain constant). We still believe that the first effect will dominate and therefore pose Hypothesis 2b:

H2b: CBF-recommenders lead to an increase in sales diversity: $G_{r_{CBF}} < G_0$

A static bestseller list promotes those products that have previously been the most popular ones. In contrast to CBF and CLF the recommendations are not individually calculated for each user, they are the same for all users. Therefore bestseller lists are not recommender systems in the classical sense. By definition the products on the bestseller list, seem to have suited the tastes of many people. In line with this there is also a high probability that these products will also be of interest to many other consumers. Due to the potentially high product fit between bestseller products and consumers, the proportion of accepted recommendations for bestseller list products is likely to be higher. This would lead to a strong focus on sales of bestseller products and therefore, compared to group 0, decrease the sales diversity as for instance shown in a simulation by Hinz and Eckert (2010). This entails Hypothesis H2c:

H2c: Bestseller lists lead to a decrease in sales diversity: $G_{r_{BS}} > G_0$

A random recommender generates recommendations independent of popularity, quality or content of the product, it just draws random products. It is therefore not a classical recommender, although the recommendations are presented as being produced by a recommender system, but the underlying data for the recommendations is just a random draw. Therefore it is pure luck if the random recommendations meet the consumer's approval and therefore, the acceptance rate of the random recommender should be lower than for other recommenders. Since the random recommender treats all tracks with an equal probability; it is quite likely that the recommendation diversity is highest among all recommenders. Therefore, even if less participants accept the random recommendations, due to the expected high diversity of the recommendations, the consequence should be a higher sales diversity. We conclude Hypothesis H2d:

H2d: Random-recommenders lead to an increase in sales diversity: $G_{r_{RD}} < G_0$

Implementation as a Web-Experiment

We believe that a web-experiment is particularly suitable for our research as it allows us to have a high level of control over the environment and thus to clearly exclude undesired effects and to measure the impact of recommender systems. In contrast to simulations and analytical models (e.g., Fleder and Hosanagar 2009) an experiment allows us to analyze actual consumer behavior. We also doubt that participants would be able to clearly isolate and state the influence of recommender systems on their purchase behavior in a survey. The usage of field data would probably not allow us to answer all of our research questions, since most providers of online shops with recommender systems would be reluctant to test different recommender technologies for different customers at the same time. In turn, we used state-of-the-art technologies and designed the experiment as an interactive web application. Our artificial "online shop" provides participants with many interactive features and is, considering its design, very similar to real online shops (see Appendix for screenshots).

The experiment is designed to model a realistic purchase environment for digital music tracks. Music seems to be very suitable for our experiment, since purchasing digital music over the Internet has become very popular among many consumers. Furthermore, music is a media product and thus considered to be an experience good. Therefore, on the one hand it is challenging to evaluate the quality of a media product prior to consumption, just based on product information. However, on the other hand consumers can quickly evaluate the quality of a music track after receiving a short trial of the product. Therefore, in our experiment, participants had the opportunity to listen to a free 30-second preview for all tracks. In addition, during the purchase process consumers in the treatment groups are supported by different types of recommender systems.

The overall selection of music tracks in the experiment was chosen based on genres with the aim to represent a genre distribution that was representative for the market in the country in which the study was conducted. For this we relied on statistics from the national chapter of the International Federation of the Phonographic Industry (IFPI). Among the different genres a random selection of old and new titles and more and less commercially successful tracks was drawn. All tracks were of the same fine digital quality.

As part of the experiment every participant had to buy 5 different music products, which they were told that every 10th participant would receive at the end of the experiment. This model is most similar to a subscription model where every customer can buy a certain number of songs per months and the recommender system has the task of supporting the customer in making quicker decisions and finding the preferred products that best match his or her taste. In order to motivate participants, they were told that every 10th participant would receive the tracks they selected during the experiment for free. Therefore, participants had an incentive to really select the songs they liked most and that they would also buy in reality. On the other hand, this also implied that participants would probably not choose the tracks they already own, even if these are their favorite tracks. However, we believe this also holds for praxis. During the experiment, participants had no time limit; therefore they could listen to music as long as they wished before conducting a purchase.

Experimental Course of Action

In a between-group post-test-only design, we implemented five different treatments that differed in the employed recommender system: a collaborative and a content-based filter as the presumably most common recommender technologies on the market, a bestseller list, a random recommender and a baseline group, with which we contrasted to all other groups. The random recommender aimed to analyze the effect of the presence of a recommender system without meaningful recommendations. It issued random recommendations, i.e. there was no underlying recommendation technology – every track was recommended with the same probability. The baseline treatment analyzed what participants purchased in our experiment environment if no recommender was present.

An overview of the structure of the experiment is presented in Figure 2. Each participant was randomly assigned to one of the five treatments. The automatically calculated recommendations were introduced as "personal recommendations" to make participants believe that the recommendations were individually calculated for them. The participants did also not learn which type of recommender technology was used in their case. To implement CBF- and CLF-recommenders, it was necessary to learn more about the participants' tastes as this served as the data basis for CBF- and CLF-recommendations. Therefore, the experiment was split into two parts: first, a rating phase in which participants were asked to evaluate music to learn more about their music taste and second, a purchasing phase, where participants received recommendations that were created by the different recommender technologies. In order to collect sufficient data for the CLF-recommendations (that were based on user similarities), we previously collected data about participants' music tastes from 32 participants (in the following they are referred to as "basic profiles"). These participants received the same experimental conditions as those in the experimental treatment without a recommender system ("none"). The similarities in the purchases of the 32 basic profiles then served as the data basis for the CLF-recommender in the experiment.

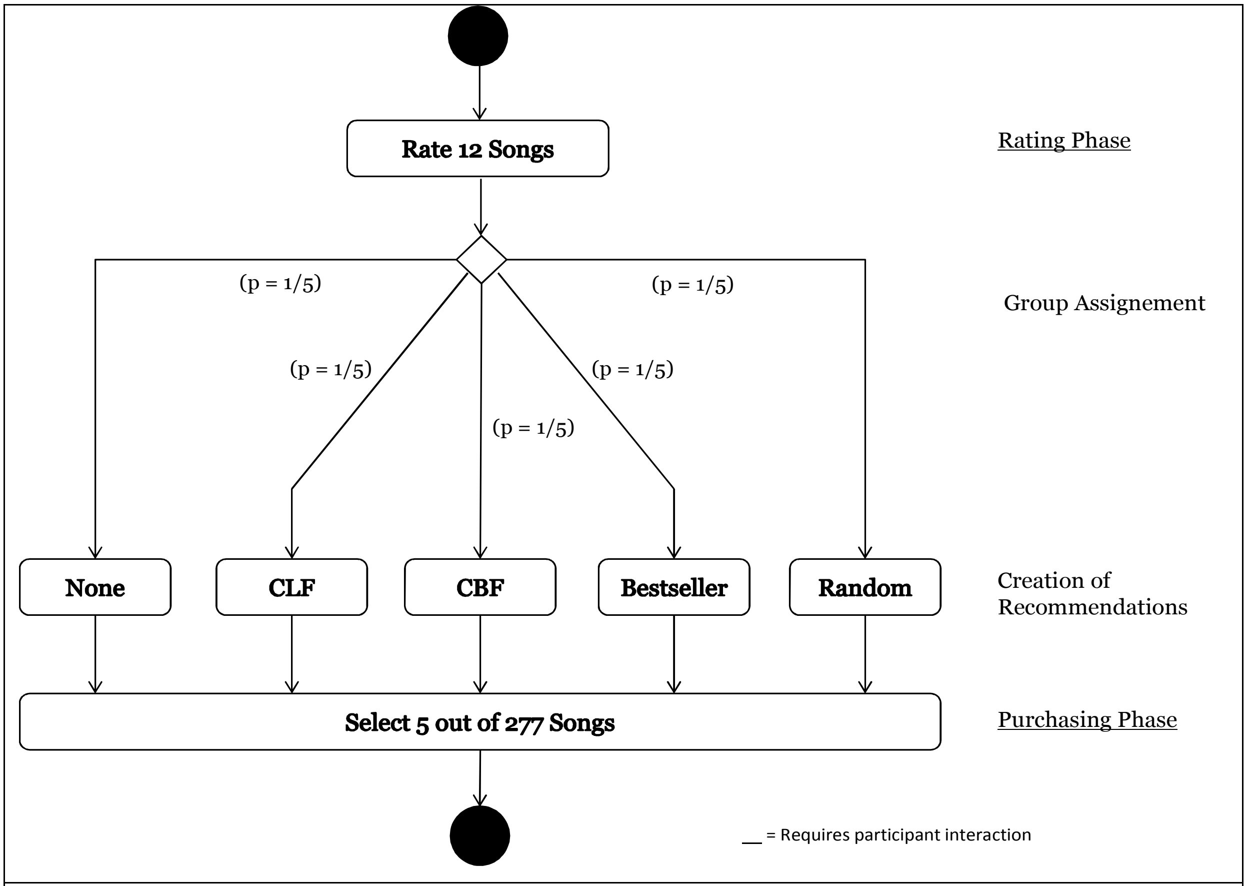


Figure 2. Structure of the Experiment

Rating Phase

In the rating phase, participants were asked to rate 12 music tracks ("rating tracks") on a Likert scale from 1 (do not like it at all) to 5 (like it very much). These tracks were selected to cover a large variety of genres, artists and ages. For complexity reasons and to ensure that participants listened to the whole selection of rating tracks, we kept this number fairly low. Still, this allowed us to get a differentiated picture of the participants' musical tastes. For the evaluation, participants could listen as many times as they wished to the 30-second samples of each track. The collected data was essential for the implementation of both

CBF- and CLF-recommenders. Furthermore, even if we had not needed any initial data for the recommenders in the purchasing phase, if we had not asked anything about their preferences before this would probably have undermined the participants' belief that recommendations were personal. Therefore the rating phase was implemented for all treatments.

Purchasing Phase

In the purchasing phase, participants were asked to select 5 out of 277 digital music tracks ("choice tracks" in the following). As in the rating phase, participants again had unlimited access to 30-second samples of each song in order to get a better idea of whether the music is to their taste. In addition, in the 4 groups with a recommender, participants received 5 recommendations of different music tracks. These tracks ("recommendation tracks") were clearly marked as recommendations and located in a dedicated part of the website. Participants were not obliged to listen to or to select any of the recommended tracks.

Implemented Recommender Technologies

Collaborative Filtering

The implemented CLF-recommender r_{CLF} was static, i.e. the underlying data was based on the basic profiles and did not get updated during the experiment. First, for each current participant p_a the most similar participant from the basic profiles was found by calculating the cosine similarity. Given p_a as the current participant with b_{p_a} as the 12-dimensional vector that contained the current participant's rating, then the cosine similarity S_{cos} to another participant from the set of basic profiles $p_i \in P$ with their respective vector b_{p_i} was:

$$S_{cos}(p_a, p_i) = \frac{b_{p_a} \cdot b_{p_i}}{\|b_{p_a}\| \, \|b_{p_i}\|}$$

The closer S_{cos} was to 1 the more similar are the two participants. Therefore, it was the highest similarity between current participant and one of the participants from the basic profiles $p_i \in P$, where the difference to 1 was minimal $S_{cos}(p_a, p_i)$. For the current participant p_a, the 5 selected tracks of the most similar participant p_i from the basic profiles were used and recommended.

Content-based Filtering

Based on all the tracks that the current participant had evaluated in the rating phase, the CBF-recommender r_{CBF} calculated similarity coefficients with all 277 tracks from the choice sample and returned those 5 tracks with the highest similarity coefficient as recommendations for the current user. For calculating the similarities between music tracks, we used a professional music comparison program that involved a number of different algorithms and parameters for the similarity calculation and that returned a matrix with content-based "distances" between all tracks. Finally, the similarity coefficient was calculated based on the degree of similarity between the two most similar songs (as reported by the music comparison program) multiplied with the participant's rating of this song. Let $t_i \in T$ be a track from the choice set T and y the most similar track from the rating set. Furthermore, let $\mathrm{R}(y)$ be the rating of the current participant p_a for the track y and $\mathrm{S}(t_i, y)$ the degree of similarity between t_i and y. Under consideration of the rating R_{p_a} of the current participant p_a the similarity coefficient Q of track t_i, was calculated as follows:

$$Q(t_i, R_{p_a}) = R(y) \cdot S(t_i, y)$$

Based on this, the participant p_a received the 5 tracks with the highest similarity coefficient as recommendations.

Bestseller List

The bestseller list r_{BS} extracted the 5 most selected tracks from the basic profiles and took these as recommendations for the current participant in the bestseller list treatment. These 5 tracks remained constant throughout the whole of the experiment. As mentioned before, the bestseller list was not a selflearning algorithm-based system and was therefore not a recommender system in the classical sense. However, due to its importance both in the physical as well as in the digital world and its simplicity to implement for shop providers, we integrated a bestseller list in our experiment.

Random Recommender

The random recommender r_{RD} selected five random song from the whole set of 277 tracks as recommendations. Therefore, each of the 277 tracks had the same probability to be drawn. Thus, the random recommender served as a pseudo control-group to analyze whether the presence of a recommender, although it had no underlying intelligence or heuristic on which the recommendations were based, had an effect on sales diversity.

Results

Descriptive Statistics

The experiment was conducted from July to September 2012 and included one pre-test to test the usability of the GUI and the readability of the task. A total of 507 participants completed the final experiment, 215 of the participants were male and 292 female. 69.4% of the participants were between 18-24 years old, another 24.7% were between 25-34. This is probably an indicator for a particularly high share of students among the participants. 72.6% of the participants claimed that they had significant experience on the Internet and in making purchases online.

In addition to demographic data and the music ratings and purchases, the number of tracks participants listened to and how much time it took participants to complete the experiment has also been recorded. Notably, the median number of played tracks is considerably lower for the baseline group without a recommender. This indicates that the recommenders were to some degree effective in encouraging participants to listen to new songs. In line with this, this is also backed up by the median number of time used to conduct the experiment, which is also higher for all treatments with recommender systems. The main characteristics for the different treatment groups are shown in Table 1.

Table 1. Overview of Participation

Recommender Type	None	CLF	CBF	Bestseller	Random
Number of Participants	102	108	100	102	95
Number of Played Tracks (Median)	27	31	31	31	32
Time Used (Median, Sec)	531	657	571	618	650

Analysis of Sales Diversity

To account for potential differences in the effects on sales diversity across recommenders, we now report the results for the different treatments. In addition to the Gini coefficients and the cumulative market share of the 95%-niche products, we also report values for the number and the distribution of the recommendations and the share of listened and purchased recommendation tracks for each of the treatments. An overview of all value is presented in Table 2.

At first sight, we see that, as already expected, the number of different recommended tracks and the resulting Gini coefficient of the recommendations varies widely. While the bestseller list was system-inherently limited to the recommendation of just five tracks, the random recommender disperses widely with the recommendation of 229 out of 277 different tracks.

The share of recommendations that participants listened to was lowest for the bestseller list (55.29% of the recommendations were played) and highest for the random recommender (68.00%). However, the picture for the purchases is just the opposite: despite having the lowest rate of recommendations played, the bestseller list has the highest share of recommendations purchased (15.88%). On the other hand, the random recommender with the highest rate of played recommendations has the lowest rate of purchases

(8.00%). This indicates that the random recommender provided participants with a lot of new songs which seemed appealing. However, many of these did not seem to suit the participants' tastes, whereas the bestseller list songs demonstrated that they are well-known by many participants again (therefore there is no reason to listen to the preview), but at the same time suit the majority's tastes.

Table 2. Main Results across Treatments

Recommender Type	None	CLF	CBF	Bestseller	Random
Recommendations					
Number of Different Recommended Songs	-	90	39	5	229
Gini Coefficient of Recommended Songs	-	.791	.930	.982	.410
Acceptance of Recommendations					
Share of Recommendations Played	-	62.22%	67.40%	55.29%	68.00%
Share of Recommendations Purchased	-	12.78%	10.00%	15.88%	8.00%
Sales Diversity					
Gini Coefficient of Sold Songs	.646	.619	.616	.645	.612
Difference to Baseline Treatment	-	-.027	-.030	-.001	-.034
Effect on Sales Diversity	-	Increase	Increase	None	Increase
Market Share of Niche Products					
Cumulated Market Share of 95%-Niche Products	71.18%	74.26%	72.20%	67.84%	73.26%
Difference to Baseline Treatment	-	+3.08%	+1.02%	-3.34%	+2.08%

The Gini coefficient G_0 of the baseline treatment group without recommender is .646, nearly equal to the Gini coefficient G_{BS} of the bestseller group with .645. In contrast, the Gini coefficients of the CLF and CBF-recommender groups G_{CLF} and G_{CBF} and also for the random recommender G_{RD} are lower. In order to test the differences for statistical significance, we used the bootstrapping method to assure the reliability of the results. For this, we used the purchase data from the experiment for each group individually and bootstrapped it with 1000 iterations. The bootstrapped Gini coefficients for all treatment groups were than compared to the baseline treatment with a two-sample t-test. The statistical tests confirm the significance of the pairwise differences between G_0 and G_{CLF}, G_{CBF} and G_{RD} ($p<.01$ for all cases). The bestseller recommender does not have a significant effect on sales diversity ($p=.073$), while all other recommenders increase the sales diversity. The pairwise differences between the groups with a recommender system were also not significant. The order of the Gini coefficients is as follows:

$$G_{RD} < G_{CBF} < G_{CLF} < G_{BS} \approx G_0$$

This serves as a first indicator that recommender systems have a positive impact on sales diversity and that the differences across technologies are rather small. To support these assumptions we now report the cumulated market share of the 95%-niche products.

The cumulated market share L_r of the 95%-niche products in the baseline treatment without a recommender is 71.18% (L_0), for the bestseller list 67.84% (L_{BS}), for the CLF-recommender 74.26% (L_{CLF}), for CBF-recommender 72.20% (L_{CBF}), and for the random recommender 73.26% (L_{RD}). Therefore, the bestseller list reduced the cumulative share of the 95%-niche products, while all other recommenders led to an increase. The order of the cumulative market share for the 95%-niche products is as follows:

$$L_{CLF} > L_{RD} > L_{CBF} > L_0 > L_{BS}$$

The additional analysis of the market share for the 95%-niche products supports the analysis of the Gini coefficients. Therefore, all recommenders except for the bestseller list have an impact on sales diversity and we can thus not reject Hypothesis H1.

Regarding the differentiating between recommender technologies, we see support for Hypotheses H2b and H2d, since both the CBF- and the random recommender increased the sales diversity as we expected. However, in contrast to our assumptions, H2b is rejected, since the CLF-recommender led to an increase in sales diversity. H2d is also rejected, since the bestseller list did not have an impact on sales diversity at all.

Discussion of the Results

Collaborative Filtering

In contrast to our assumptions, the CLF-recommender did not lead to a decrease in sales diversity; rather it actually increased the sales diversity and also the share of 95%-niche products. A potential reason for this is the concrete implementation of the CLF-recommender that was used in our experiment. The recommender was designed as a static system, i.e. all recommendations were based on the set of 32 basic user profiles that were initially collected. This allowed us to overcome the cold-start-problem, which states that only products that are already purchased or rated can be recommended (Schein et al. 2002). Thus, the diversity of the recommended products from the beginning of our experiment may have been higher than in praxis in our case. However, also in praxis, providers of online shops suffer from little data after a new product is introduced to the market and therefore proxy data or other recommendation methods are frequently used to overcome this problem. Furthermore, since the data basis remained static during the experiment, this may have also decreased the self-enhancing effect of the CLF-recommender, by which already popular products become even more popular (Fleder and Hosanagar 2009).

Content-based Filtering

As expected the CBF-recommender increased the sales diversity in our experiment. Since CBF-recommenders take only product characteristics, but not the popularity of the product into account, they may recommend products that have not been purchased or rated before. Therefore, especially niche products, that may otherwise be less likely to be purchased or rated and thus recommended by a CLF-recommender, can be recommended by CBF-recommenders. However, the share of recommended products which were purchased was lower for the CBF-recommender than for the CLF-recommender, which may confirm the assumption that CBF-recommenders are not as suitable as CLF-recommenders for media products.

The fairly low recommendation diversity in the experiment was to some degree probably system-inherent, since the recommender suggested songs that are similar to the primarily evaluated tracks in the rating phase. Therefore the number of tracks that could be recommended during the purchasing phase was limited to the 5 most similar tracks to any of the 12 primarily evaluated rating tracks and therefore has a natural maximum bound of 60 in our case. Out of these 60 tracks, 39 different tracks were finally recommended. Still, the CBF-recommender increased the sales diversity as we had already expected.

Bestseller List

Despite the fact that the bestseller list suggested only 5 common tracks to all participants and had therefore the smallest recommendation diversity, the Gini value for the sales diversity compared to the baseline treatment apparently remained almost constant. However, it is worth mentioning that although the bestseller-recommender did not lead to a change in the Gini coefficient, it led to the largest change in the cumulative market share of the 95%-niche products. In this context, a visual inspection of the Lorenz curve is interesting and leads to additional insights (cp. Figure 3). As already mentioned before, it is possible that different Lorenz-curves have the same Gini coefficient. In this case, the Lorenz curve illustrates that the distinct reduction in the cumulative market share of the 95%-niche products only had no effect on the Gini coefficient, since it was balanced out by the simultaneous increase of products in the 75% cumulative market share percentile range. Therefore, it was no pure shift from niche to bestseller products, but rather a flattening of the Lorenz curve.

One reason for the high share of recommended songs that were purchased by participants is that these tracks probably suit the tastes of a large number of participants. In addition, some of these tracks may have been quite popular (e.g., in the charts and receiving a great deal of airtime) when the experiment was conducted and therefore, due to this, a higher number of participants may have been willing to purchase these tracks anyway. The comparatively low rate of the recommended tracks that were played by participants was probably also caused by the high-profile on radio etc. of these songs at the time of the experiment.

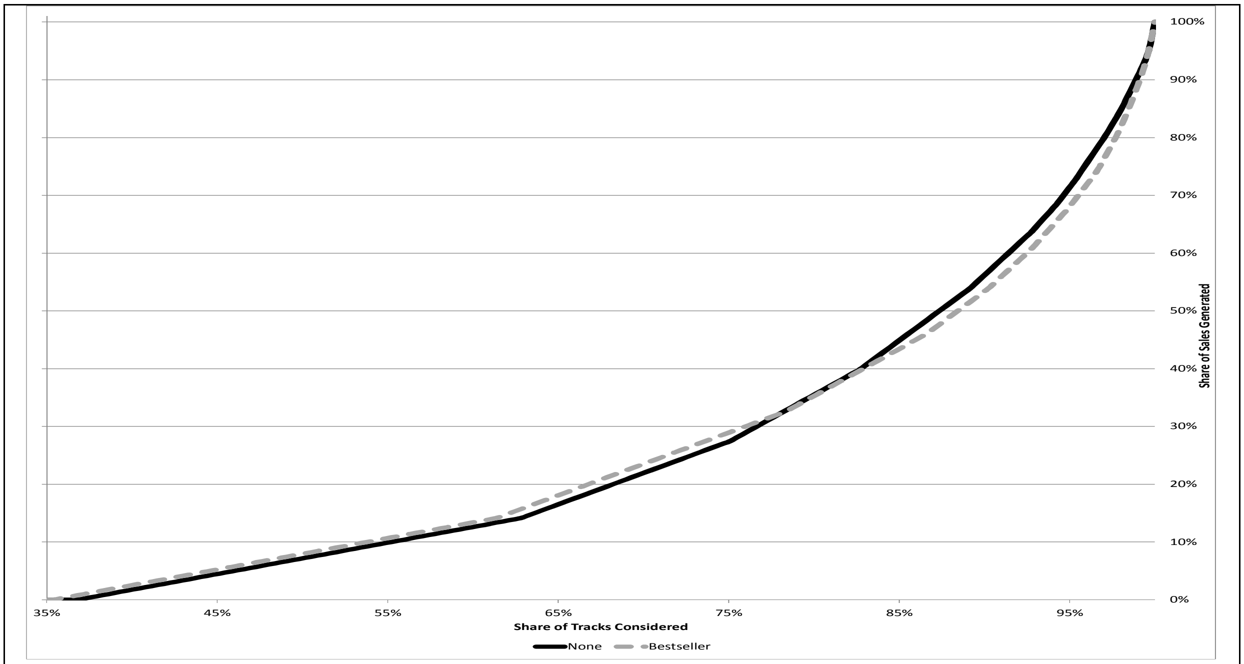


Figure 3. Lorenz Curves for Bestseller List and Baseline Treatment

Random Recommender

Although most participants listened to songs that were recommended by the random recommender, the share of purchased recommendations was lower than for all other recommenders. Therefore, the random recommender succeeded in providing new songs to participants, but these recommendations mostly did not meet participants' tastes. However, it is probably due to the high diversity of recommendations that, despite the low acceptance rate of recommendations, the increase in sales diversity was highest among all treatments.

Summary and Further Implications

The purpose of this study was to analyze the influence of different recommender system technologies on sales diversity. For this, we used a realistic web-based experiment, in which different groups of participants supported by different recommender technologies were asked to purchase a number of digital music tracks. We also implemented a baseline group, where no recommender system was present in order to monitor participant behavior without recommendations and to have a reference value for comparing the influence of the different recommender technologies. For measuring potential differences in sales diversity we used the Gini coefficient and in addition the cumulative market share of the 95%-niche products. The statistical tests for potential differences used the bootstrapping method to get reliable results.

Our findings have interesting implications for the ongoing fragmentation debate, since in contrast to most previous studies we analyzed changes in the sales distribution for different recommender technologies. Our results indicate that in comparison to the scenario where no recommender was present, Collaborative-Filtering and Content-Based-Filtering increase sales diversity, i.e. that on average consumers tend to buy quite diverse products, which in turn fosters markets for long-tail products. However, a random recommender system also led to an increase in sales diversity, while the extent of the decrease was similar when compared to the former two recommender technologies. In contrast to this, only the bestseller list had no effect on sales diversity. These findings were supported both by the Gini coefficient as well the cumulative market share of the 95%-niche products, except that for the bestseller list the share of niche products even decreased. Apparently, the first three recommender types successfully propose new products to consumers, which they would otherwise not find, while bestsellers mostly propose products that are per se more likely to be well-known and chosen, or that have already been purchased by a large number of consumers.

Therefore, we conclude that the type of recommender system does have an impact, not just on the individual purchase process of a single consumer, but also on the aggregated sales diversity. For providers of online shops this means that those who have an interest in selling high volumes of a small number of different products should use and highlight bestseller lists. Due to the similarities between the Gini coefficient and Lorenz curve when compared to the scenario where no recommender was used, it would also be possible to refrain from using any form of recommender system. However, we would not advise this since this might be seen as outdated by consumers. Retailers would also not make use of recommender systems' potential to reduce information overload and to increase sales.

On the other hand, online shop providers who actively want to promote their extensive product breadth are well-advised to use any of the other recommender system, since all increase sales diversity to a similar extent. However, using a random recommender could be seen as cheating to some extent ("personal recommendations") and these systems also suffer from a lower consumer acceptance. Due to the constituted impact of the recommender technology on the sales diversity, the choice of a concrete recommender technology may also have important implication for suppliers of products. If many large online shops decide to use recommender technologies that increase the sales diversity, this has positive effects on suppliers of niche products, while the use of bestseller lists would be more suited to large media companies which focus on the production of blockbuster products.

Limitations of the Study

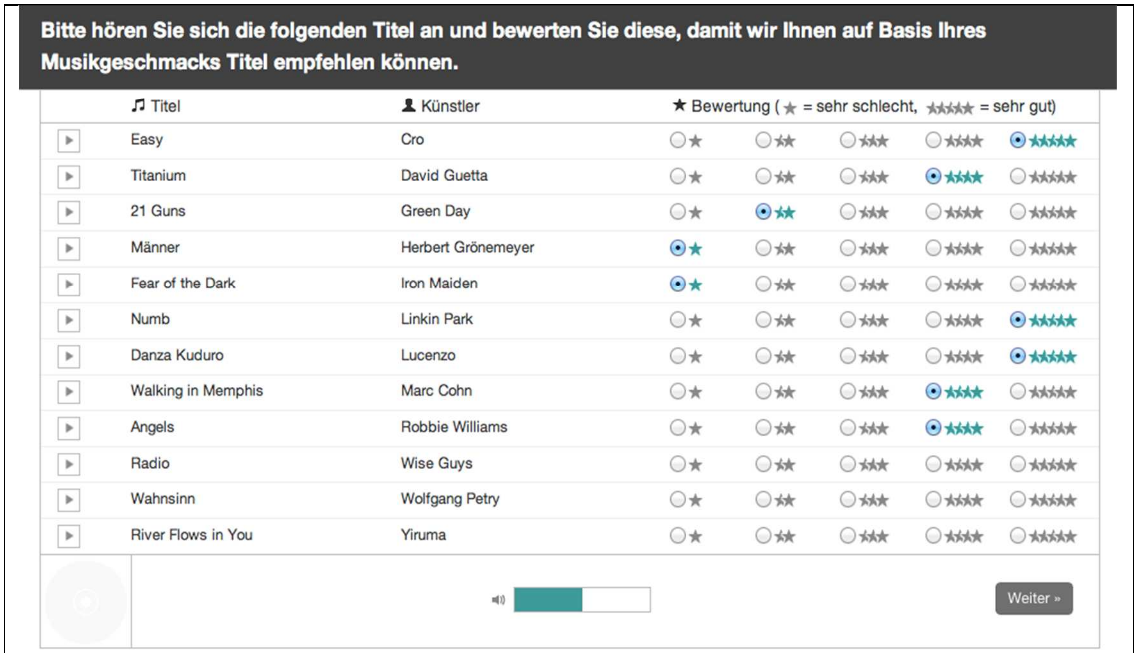
Despite extensive efforts to provide a realistic experimental environment, the current study at hand is subject to several limitations. First of all, our experiment focused on digital music tracks only. As a form of experience goods, music possesses distinct characteristics, i.e. search goods are usually easier to describe and to evaluate their quality prior to a purchase. The effect of recommender systems may thus be different for experience goods. This study should therefore be replicated with other products.

Secondly, as previously mentioned, the sample consisted of a disproportionately high proportion of students. We believe that due to our incentivization approach and the fact that students are a very active buyer group online, they are appropriate for our study. However, potential variations to a representative sample of Internet users should be tested.

Thirdly, for complexity reasons and due to the comparably short duration of the experiment we used a static CLF-recommender, i.e. the underlying data consisted of a limited number of profiles that were initially set-up and not updated during the experiment. However, in praxis most recommender systems are updated with new data when new users sign up or through other events, such as when purchases or product ratings are conducted (Linden et al. 2003; Resnick et al. 1994). Furthermore, there is a large variety of specific forms of recommender systems on the market. Some online shops also use hybrid systems, combining several basic recommender technologies. Obviously, due to the applied methodology we were not able to cover all possible forms of recommender technology and, for the sake of simplicity, focused on the main basic methods.

Finally, to ensure a balanced design and to make participant-behavior comparable, we set the number of tracks participants were asked to buy to five for everyone. In praxis, some participants would probably have bought more tracks whereas others would be likely to purchase less. In addition, this restriction limited our possibilities to state whether the provision of certain recommender system types would lead to changes in the number of total sales of music tracks.

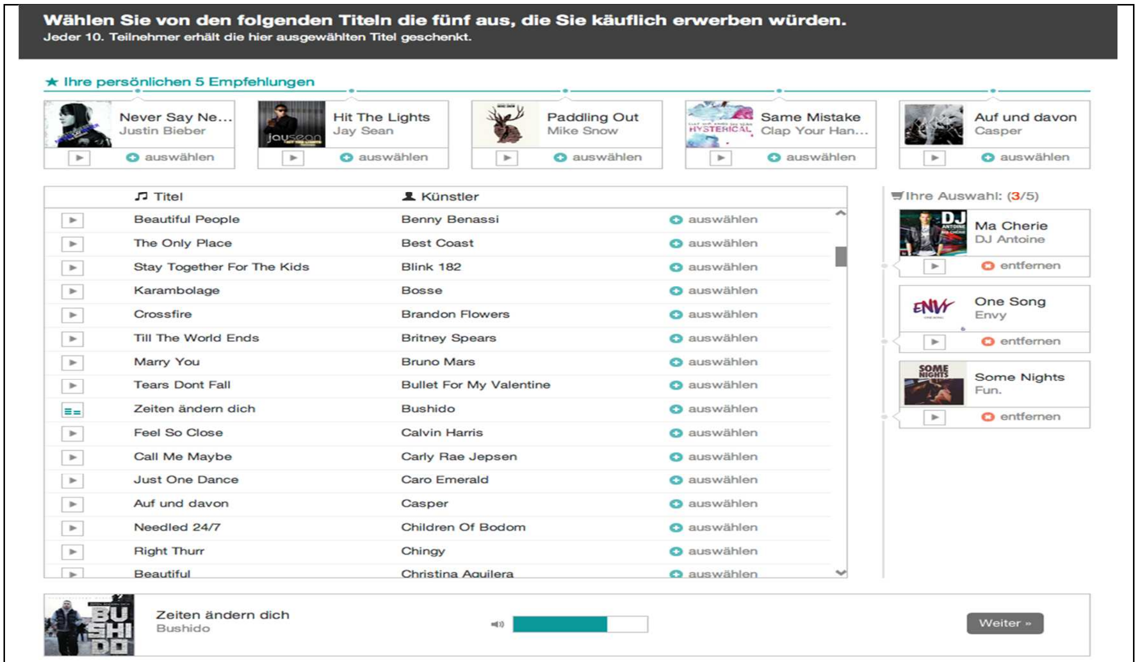
Appendix



Bitte hören Sie sich die folgenden Titel an und bewerten Sie diese, damit wir Ihnen auf Basis Ihres Musikgeschmacks Titel empfehlen können.

	♫ Titel	Künstler	★ Bewertung (★ = sehr schlecht, ★★★★★ = sehr gut)				
▶	Easy	Cro	○ ★	○ ★★	○ ★★★	○ ★★★★	● ★★★★★
▶	Titanium	David Guetta	○ ★	○ ★★	○ ★★★	● ★★★★	○ ★★★★★
▶	21 Guns	Green Day	○ ★	● ★★	○ ★★★	○ ★★★★	○ ★★★★★
▶	Männer	Herbert Grönemeyer	● ★	○ ★★	○ ★★★	○ ★★★★	○ ★★★★★
▶	Fear of the Dark	Iron Maiden	● ★	○ ★★	○ ★★★	○ ★★★★	○ ★★★★★
▶	Numb	Linkin Park	○ ★	○ ★★	○ ★★★	○ ★★★★	● ★★★★★
▶	Danza Kuduro	Lucenzo	○ ★	○ ★★	○ ★★★	○ ★★★★	● ★★★★★
▶	Walking in Memphis	Marc Cohn	○ ★	○ ★★	○ ★★★	● ★★★★	○ ★★★★★
▶	Angels	Robbie Williams	○ ★	○ ★★	○ ★★★	● ★★★★	○ ★★★★★
▶	Radio	Wise Guys	○ ★	○ ★★	○ ★★★	○ ★★★★	○ ★★★★★
▶	Wahnsinn	Wolfgang Petry	○ ★	○ ★★	○ ★★★	○ ★★★★	○ ★★★★★
▶	River Flows in You	Yiruma	○ ★	○ ★★	○ ★★★	○ ★★★★	○ ★★★★★

Figure 4. Screenshot of Rating Phase



Wählen Sie von den folgenden Titeln die fünf aus, die Sie käuflich erwerben würden.
Jeder 10. Teilnehmer erhält die hier ausgewählten Titel geschenkt.

★ Ihre persönlichen 5 Empfehlungen

Titel	Künstler	
Never Say Ne…	Justin Bieber	▶ ⊕ auswählen
Hit The Lights	Jay Sean	▶ ⊕ auswählen
Paddling Out	Mike Snow	▶ ⊕ auswählen
Same Mistake	Clap Your Han…	▶ ⊕ auswählen
Auf und davon	Casper	▶ ⊕ auswählen

	♫ Titel	Künstler	
▶	Beautiful People	Benny Benassi	⊕ auswählen
▶	The Only Place	Best Coast	⊕ auswählen
▶	Stay Together For The Kids	Blink 182	⊕ auswählen
▶	Karambolage	Bosse	⊕ auswählen
▶	Crossfire	Brandon Flowers	⊕ auswählen
▶	Till The World Ends	Britney Spears	⊕ auswählen
▶	Marry You	Bruno Mars	⊕ auswählen
▶	Tears Dont Fall	Bullet For My Valentine	⊕ auswählen
≡	Zeiten ändern dich	Bushido	⊕ auswählen
▶	Feel So Close	Calvin Harris	⊕ auswählen
▶	Call Me Maybe	Carly Rae Jepsen	⊕ auswählen
▶	Just One Dance	Caro Emerald	⊕ auswählen
▶	Auf und davon	Casper	⊕ auswählen
▶	Needled 24/7	Children Of Bodom	⊕ auswählen
▶	Right Thurr	Chingy	⊕ auswählen
▶	Beautiful	Christina Aguilera	⊕ auswählen

Ihre Auswahl: (3/5)

Titel	Künstler	
Ma Cherie	DJ Antoine	▶ ⊗ entfernen
One Song	Envy	▶ ⊗ entfernen
Some Nights	Fun.	▶ ⊗ entfernen

Zeiten ändern dich
Bushido

Figure 5. Screenshot of Purchasing Phase

References

Adomavicius, G., and Tuzhilin, A. 2005. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering* (17:6), pp. 734-749.

Anderson, C. 2009. *The Long Tail: How Endless Choice Is Creating Unlimited Demand*. London: Random House.

Apple. 2013. "*Itunes Store Sets New Record with 25 Billion Songs Sold*." from http://www.apple.com/pr/library/2013/02/06iTunes-Store-Sets-New-Record-with-25-Billion-Songs-Sold.html

Balabanović, M., and Shoham, Y. 1997. "Fab: Content-Based, Collaborative Recommendation," *Communications of the ACM* (40:3), pp. 66-72.

Basu, C., Hirsh, H., and Cohen, W. 1998. "Recommendation as Classification: Using Social and Content-Based Information in Recommendation," in *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI 98)*, Madison, WI, USA, pp. 714-720.

Breese, J. S., Heckerman, D., and Kadie, C. 1998. "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI '98)*, San Francisco, CA, USA, pp. 43-52.

Brynjolfsson, E., Hu, Y. J., and Simester, D. 2011. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," *Management Science* (57:8), pp. 1373-1386.

Brynjolfsson, E., Hu, Y. J., and Smith, M. D. 2006. "From Niches to Riches: Anatomy of the Long Tail," *MIT Sloan Management Review* (47:4), pp. 67-71.

Brynjolfsson, E., Hu, Y. J., and Smith, M. D. 2010. "Long Tails Vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns," *Information Systems Research* (21:4), pp. 736-747.

Burke, R. 2000. "Knowledge-Based Recommender Systems," in *Encyclopedia of Library and Information Systems, Vol. 69, Supplement 32*, A. Kent (ed.). New York: Dekker, pp. 175-186.

Burke, R. 2002. "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction* (12:4), pp. 331-370.

Dias, M. B., Locher, D., Li, M., El-Deredy, W., and Lisboa, P. J. G. 2008. "The Value of Personalised Recommender Systems to E-Business: A Case Study," in *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys '08)*, New York, NY, USA, pp. 291-294.

Fleder, D., Hosanagar, K., Lee, D., and Buja, A. 2012. "Recommender Systems and Their Effects on Consumers: The Fragmentation Debate," NET Institute Working Paper No. 08-44, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1321962.

Fleder, D. M., and Hosanagar, K. 2009. "Recommender Systems and Their Impact on Sales Diversity," *Management Science* (55:5), pp. 697-712.

Gini, C. 1912. *Variabilità E Mutabilità: Contributo Allo Studio Delle Distribuzioni E Delle Relazioni Statistiche*. Cagliari, Italy: Facoltá di Giurisprudenza della R. Universitá dei Cagliari.

Hinz, O., and Eckert, J. 2010. "The Impact of Search and Recommendation Systems on Sales in Electronic Commerce," *Business & Information Systems Engineering* (2:2), pp. 67-77.

Hinz, O., Eckert, J., and Skiera, B. 2011. "Drivers of the Long Tail Phenomenon: An Empirical Analysis," *Journal of Management Information Systems* (27:4), pp. 43-70.

Jannach, D., and Hegelich, K. 2009. "A Case Study on the Effectiveness of Recommendations in the Mobile Internet," in *Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09)*, New York, NY, USA, pp. 205-208.

Komiak, S. Y. X., and Benbasat, I. 2006. "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," *MIS Quarterly* (30:4), pp. 941-960.

Koren, Y., Bell, R., and Volinsky, C. 2009. "Matrix Factorization Techniques for Recommender Systems," *Computer* (42:8), pp. 30-37.

Kumar, N., and Benbasat, I. 2006. "Research Note: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites," *Information Systems Research* (17:4), pp. 425-439.

Linden, G., Smith, B., and York, J. 2003. "Amazon.Com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing* (7:1), pp. 76-80.

Mooney, R. J., and Roy, L. 2000. "Content-Based Book Recommending Using Learning for Text

Categorization," in *Proceedings of the Fifth ACM Conference on Digital Libraries*, San Antonio, TX, USA, pp. 195-204.

Oestreicher-Singer, G., and Sundararajan, A. 2012. "The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets," *Management Science* (58:11), pp. 1963-1981.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. 1994. "Grouplens: An Open Architecture for Collaborative Filtering of Netnews," in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, Chapel Hill, NC, USA, pp. 175-186.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2000. "Analysis of Recommendation Algorithms for E-Commerce," in *Proceedings of the Second ACM Conference on Electronic Commerce*, Minneapolis, MN, USA, pp. 158-167.

Schafer, J. B., Konstan, J., and Riedi, J. 1999. "Recommender Systems in E-Commerce," in *Proceedings of the 1st ACM Conference on Electronic Commerce*, Denver, CO, USA, pp. 158-166.

Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M. 2002. "Methods and Metrics for Cold-Start Recommendations," in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*, New York, NY, USA, pp. 253-260.

Senecal, S., and Nantel, J. 2004. "The Influence of Online Product Recommendations on Consumers' Online Choices," *Journal of Retailing* (80:2), pp. 159-169.

Shani, G., Brafman, R. I., and Heckerman, D. 2005. "An Mdp-Based Recommender System," *The Journal of Machine Learning Research* (6:2005), pp. 1265-1295.

Szlavik, Z., Kowalczyk, W., and Schut, M. C. 2011. "Diversity Measurement of Recommender Systems under Different User Choice Models," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain.

Tucker, C., and Zhang, J. 2011. "How Does Popularity Information Affect Choices? A Field Experiment," *Management Science* (57:5), pp. 828-842.

Wang, J., De Vries, A. P., and Reinders, M. J. T. 2006. "Unifying User-Based and Item-Based Collaborative Filtering Approaches by Similarity Fusion," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*, New York, NY, USA, pp. 501-508.

Xiao, B., and Benbasat, I. 2007. "E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact," *MIS Quarterly* (31:1), pp. 137-209.