Serendipity in the city: user evaluations of urban recommender systems

Smets, Annelien; Vannieuwenhuyze, Jorre; Ballon, Pieter

Published in:
Journal of The Association For Information Science And Technology

DOI:
10.1002/asi.24552

Publication date:
2021

Document Version:
Accepted author manuscript

Serendipity in the city: user evaluations of urban recommender systems

Annelien Smets*, Jorre Vannieuwenhuyze, and Pieter Ballon

imec-SMIT, Vrije Universiteit Brussel

Final draft. To cite this article: Smets, A., Vannieuwenhuyze, J., & Ballon, P. (2022). Serendipity in the city: User evaluations of urban recommender systems. Journal of the Association for Information Science and Technology, 73(1), 19-30. https://doi.org/10.1002/asi.24552

Abstract: The contemporary city is increasingly being labeled as a smart city consisting of both physical and virtual spaces. This digital augmentation of urban life sets the scene for urban recommender systems to help citizens dealing with the abundance of digital information and corresponding choice overload, for example, by recommending the best place to have dinner based on your personal profile. There are, however, concerns that this kind of algorithmic filtering could lead to homogenization of urban experiences and a decline of social cohesion among citizens. To overcome this issue, scholars increasingly encourage the introduction of serendipity in all types of recommender systems. Nonetheless, it remains unclear how this can be achieved in practice. In this work, we study user evaluations of serendipity in urban recommender systems through a survey among 1641 citizens. More specifically, we study which characteristics of recommended items contribute to serendipitous experiences and to what extent this increases user satisfaction and conversion. Our results align with findings in other application domains in the sense that there is a strong relation between the relevance and novelty of recommendations and the corresponding experienced serendipity. Moreover, serendipitous recommendations are found to increase the chance of users following up on these recommendations.

Keywords: Filter Bubble; Serendipity; Smart City; Urban Recommender Systems; User Study

*Corresponding author: imec-SMIT, Vrije Universiteit Brussel, Pleinlaan 9, 1050 Brussels, Belgium, `annelien.smets@vub.be`

1 Introduction

In contemporary cities, activities and experiences are increasingly shaped by digital urban media (de Waal, 2013) such as urban recommender systems (Quijano-Sánchez, Cantador, Cortés-Cediel, & Gil, 2020). Citizens use these systems to help choose among possible destinations and activities (e.g. TripAdvisor), places to eat or rest (e.g. Airbnb) or navigating through the urban environment (e.g. Waze). These urban recommender systems are an important example of real-time merging of digital and physical public space to achieve efficiency and productivity. As such, they are argued to contribute to the so-called 'smart city' (van der Graaf & Ballon, 2019).

Although urban recommender systems help citizens to cope with the abundance of available digital information, they are often built on the premise that users are merely looking for information closely matching their own profiles (Iaquinta et al., 2008; Jannach & Adomavicius, 2016). For example, recently it has been shown that Google Maps guides people to the very same types of restaurants, museums or activities over and over again based on their demographic profiles and search history (Smets, Montero, & Ballon, 2019). As a consequence of this premise, concerns are raised that recommender systems may create urban 'filter bubbles' (see Pariser, 2011) by merely exposing users to predictable, popular and homogeneous content rather than challenging their views with serendipitous encounters (Zuiderveen Borgesius et al., 2016; Foth, 2017).

Serendipity refers to "what happens when we, in unplanned ways, encounter resources that we find interesting" (Björneborn, 2017, p.2). As such, it has been argued that serendipitous encounters in the city are not only key factors for cities' economic and innovative growth (Wood & Landry, 2008), but that they are also central to social bonding and urban trust as described by Jacobs' (1961) notion of 'well-used streets'. As a result, urban recommender systems ignoring serendipitous recommendations may form a threat to the urban ecosystem (McQuire, 2017). This threat of people not being exposed to and excluded from spontaneous interaction with the diversity of cities and their inhabitants is a timely issue that needs urgent attention (Smets et al., 2019).

Recently, there has been an increased focus on serendipity in the domain of recommender systems. Nonetheless, this line of research still shows some gaps. Firstly, serendipity in recommender systems has rarely been investigated within an urban context. Over the last decade, the use of recommender systems and their failure to introduce serendipity has mainly been discussed in the context of mere online (social media) platforms (Reviglio, 2019). Only a few preliminary urban digital information systems have been developed introducing serendipity into urban recommender systems, as a merger between a digital and a physical environment rather than an entirely digital environment. These few examples include applications for urban navigation (X. Ge et al., 2017; Delva, Smets, Colpaert, Ballon, & Verborgh, 2020; Li & Tuzhilin, 2019; Shepard, 2011) or applications to connect strangers in public places (Paulos & Goodman, 2004). Research on serendipity has, however, demonstrated that serendipity evolves differently in different contexts inducing the need to further investigate serendipity in urban recommender systems in particular (Lutz, Pieter Hoffmann, & Meckel, 2017; Sun, Sharples, & Makri, 2011; Olshannikova, Olsson, Huhtamäki, Paasovaara, & Kärkkäinen, 2020).

Secondly, the small amount of existing studies about serendipity in urban recommender systems rarely focus on user experience and evaluation in the field. The majority of these studies rely on offline lab experiments for design optimization (Kotkov, Veijalainen, & Wang, 2016). Such experiments can be useful in choosing candidate algorithms, but often take place in artificial settings in which specific recommender systems are imposed on test subjects. In order to circumvent the limitations of such lab experiments, large scale field studies are required to investigate user feedback in daily life situations (Kotkov, Veijalainen, & Wang, 2016). Due to the involvement of actual users those field studies are costlier and therefore often neglected (Silveira, Zhang, Lin, Liu, & Ma, 2019).

This study aims to fill these gaps by investigating user evaluations of serendipity in urban recommender systems. More specific, this study addresses the question which characteristics of urban recommendations lead to serendipity experiences and to what extent this increases user satisfaction and conversion (i.e. the capacity of the recommender system to convince users to follow up on the recommendations). In this way, the results of this study provide a first insight into user evaluations of serendipity in urban recommender systems.

The paper is structured as follows. The next section provides an overview of the literature on urban recommender systems and the role of serendipity in those systems. This overview will lead to a set of research questions under study in this work. Section 3 introduces a survey among 1641 citizens about experienced serendipity, user satisfaction and conversion in urban contexts.

Section 4, subsequently, elaborates on the results of the analyzed survey data. Finally, section 5 ends the paper with a discussion of the results.

2 Theoretical framework

2.1 Serendipity

Serendipity in digital environments has been studied from a wide range of perspectives, each emphasizing their own focus and assumptions (Reviglio, 2019). In information science and information behavior research, emphasis is usually put on process models explaining the occurrence of serendipity experiences (Erdelez & Makri, 2020; Lutz et al., 2017; Erdelez, 2004; Makri & Blandford, 2012). Various authors, for example, identified several contextual factors related to the user, information, tasks and the environment that influence the process of information encountering (e.g. Jiang, Liu, & Chi, 2015; Erdelez & Makri, 2020). Such contextual factors are also adopted in the field of information system design, in which they are considered precipitating conditions increasing the likelihood of serendipity (McCay-Peet & Toms, 2011; Björneborn, 2017). Several studies have been conducted trying to assess these conditions and the extent to which they contribute to serendipity. Here, however, a mere distinction is usually made between personal cognitive and behavioral antecedents (e.g. Lutz et al., 2017) versus environmental factors or characteristics (e.g. McCay-Peet & Toms, 2011). Those environmental factors consist of various aspects that have been categorized by Björneborn (2017) as three key affordances: diversifiability, traversability and sensoriability. These affordances should be considered as building blocks when designing environments that facilitate serendipity, and respectively refer to the ability of the environment to allow a diversity of contents, to be traversable and to be perceivable by the senses.

In the context of recommender systems, environmental affordances refer to characteristics of the system itself. That is the diversity of the recommended items, the navigation and interactivity of the system and the user interface design. Most of the work on serendipity in recommender systems has, however, mainly been dealing with the recommended items themselves. In this strand of the literature, serendipity is most commonly considered as a compound concept consisting of three characteristics of recommended items: relevance, novelty and diversity (Chen, Yang, Wang, Yang, & Yuan, 2019; Kotkov, Wang, & Veijalainen, 2016; also see Figure 1).

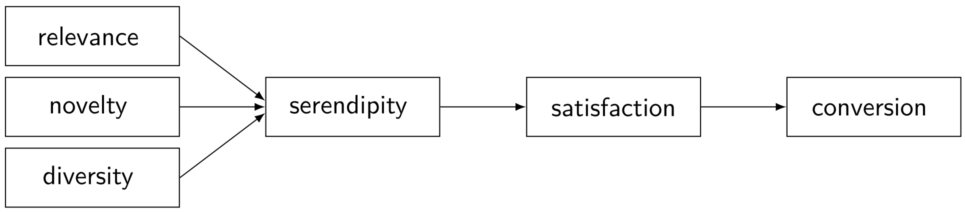


Figure 1. We hypothesize that the relevance, novelty and diversity of recommended items in urban recommender systems affect user satisfaction and, subsequently, user conversion through experienced serendipity.

Relevance refers to recommended items that users like or are interested in (Maksai, Garcin, & Faltings, 2015; Iaquinta et al., 2008; Kotkov, Wang, & Veijalainen, 2016) and is usually measured by the accuracy of predicted user ratings for unseen items. It is an important aspect of serendipitous recommendations because, by definition, serendipity refers to encounters that are relevant to the user. Nonetheless, sole focus on relevance may lead to filter bubbles because it precludes unexpectedness (Kotkov, Wang, & Veijalainen, 2016).

Indeed, serendipity also requires unexpectedness, which may be facilitated by including *novelty* and *diversity* in recommended items (Kotkov, Veijalainen, & Wang, 2016; Tacchini, 2012; M. Ge, Delgado-Battenfeld, & Jannach, 2010). *Novelty* refers to items that are unknown to the user either because they are (1) novel to the system, (2) forgotten by the user, (3) unknown to the user or (4) unrated by the user (Kotkov, Wang, & Veijalainen, 2016; Iaquinta et al., 2008). *Diversity* refers to the variability in recommended items a user receives (Kotkov, Wang, & Veijalainen, 2016). It was already found by Chen et al. (2019) that both novelty and diversity are important antecedents of unexpectedness, which in turn, affects experiences of serendipity. Nonetheless, the overall influence of diversity on serendipity was not confirmed by their study. The availability of a diverse set of items is, however, considered to be a key environmental affordance to foster serendipity (Björneborn, 2017).

The hypothesized influence of relevance, novelty and diversity on experienced serendipity brings us to our first research question:

> **RQ1:** How do relevance, novelty and diversity affect users' experienced serendipity in urban recommender systems?

2.2 User satisfaction and conversion

The second goal of this study is to investigate whether experiences of serendipity in urban recommender systems also lead to higher user satisfaction and, consequently, higher user conversion rates (i.e. the ability of the recommender system to persuade users to actually follow up on the recommendations). After all, it has been assumed that sole focus on relevance, in contrast to serendipity, does not optimize user satisfaction because users do not appreciate lists with very similar items (Kotkov, Veijalainen, & Wang, 2016; De Gemmis, Lops, Semeraro, & Musto, 2015; Chen et al., 2019; Lutz et al., 2017; Zhang, Séaghdha, Quercia, & Jambor, 2012; Said, Fields, Jain, & Albayrak, 2013). User satisfaction, in turn, has been shown to increase user conversion (Chen et al., 2019; Venkatesh, Thong, & Xu, 2012).

Within the existing literature, it has already been shown that diversity positively correlates with user satisfaction (Kotkov, Veijalainen, & Wang, 2016; De Gemmis et al., 2015; Kunaver & Požrl, 2017). In the context of serendipity, however, these findings contradict with the results of previously mentioned studies where only small relations were found between diversity and experiences of serendipity (Chen et al., 2019).

Mixed results have been found for the relationship between novelty and user satisfaction (Ekstrand, Harper, Willemsen, & Konstan, 2014; Chen et al., 2019). Novel items do not necessarily positively correlate with user satisfaction or conversion because novelty could also decrease users' trust in the capabilities of the system (Ekstrand et al., 2014).

However, given the call for contextual differentiation in serendipity research and the current limited focus on serendipity in urban recommender systems, this study aims to further investigate the relation between serendipity antecedents and user satisfaction. In sum, we hypothesize that relevance, novelty and diversity affect user satisfaction and, subsequently, user conversion through experienced serendipity (see Figure 1). This brings us to the following research questions:

> **RQ2:** Do relevance, novelty and diversity influence users' satisfaction in urban recommender systems and can this influence be explained by experienced serendipity?
>
> **RQ3:** Do relevance, novelty and diversity influence user conversion in urban recommender systems and can this influence

be explained by experienced serendipity and user satisfaction?

2.3 Contextual differentiation

The third goal of this study is to investigate whether experienced serendipity also depends on the system's domain or user's needs. After all, it has already been shown that users might have different needs for serendipity in different recommendation scenarios (McCay-Peet, 2014; Kaminskas & Bridge, 2016; Ekstrand et al., 2014; Sun et al., 2011). Indeed, serendipity in recommender systems has already been studied in various contexts such as e-commerce (Chen et al., 2019; Lutz et al., 2017), movies (Kotkov, Konstan, Zhao, & Veijalainen, 2018), music (Zhang et al., 2012; Matt, Benlian, Hess, & Weiß, 2014) and social networking sites (Lutz et al., 2017), all leading to variable results. As a result, acknowledging that the urban environment is eminently heterogeneous, studying serendipity in urban recommender systems also requires to take a contextual differentiation into account. This brings us to our final research question:

RQ4: Does the impact relevance, novelty and diversity on experienced serendipity in urban recommender systems depend on the context of use?

3 Methods

3.1 Sample

Most existing studies on serendipity experiences with recommender systems adopt an experimental approach in which test persons are confronted with an artificial recommender system after which they are immediately asked for their experiences (for example Pu, Chen, & Hu, 2011; Ekstrand et al., 2014). However, especially in the context of offline activities, such experimental studies create settings that may significantly deviate from regular situations in daily life in which people use recommender systems. As an alternative, we decided to use a survey with retrospective questions about actual behavior and experiences in real-life situations.

We investigated the user evaluations of serendipity in urban recommender systems through the Smart City Meter 2020. This is an annual survey in Flanders and Brussels, Belgium, about citizens' opinions, attitudes and behaviors in the context of smart cities. The data were collected between

March 1 and April 30, 2020 among people recruited through an online panel of a private market research agency. This panel had been collected and maintained over the years through various projects of the agency. A stratified sample was taken from this panel according to gender, age and place of residence (Brussels, Antwerp, Ghent, other large cities, small towns, municipalities). The size of the strata was determined by the distribution of these variables among the Brussels and Flemish population. In addition, strata size was adjusted according to the response rate within each stratum based on previous experience of the agency. However, all drawn panel members were invited by email to complete the questionnaire in the same way and with the same number of contact attempts.

In total, 1641 eligible panel members responded to the questionnaire. Because of the disproportional stratified sampling strategy, the realized sample more or less followed the population distribution of gender, age and place of residence. Nonetheless, because of panel constraints, people under the age of 30 and over the age of 70 were slightly underrepresented. For that reason, based on the population distributions, the respondents were assigned analysis weights. Further analysis also showed that the sample included both higher and lower educated people and voters of all relevant political parties.

3.2 Variables

In order to allow for contextual differentiation (cf. *RQ4*) the respondents were randomly divided into two groups. The first group was confronted with questions about recommender systems for Catering in the city (restaurants and bars). The second group, in turn, was confronted with questions about recommender systems for general Activities in the city. The respondents were firstly asked how often they use recommender systems (websites or apps like Google, TripAdvisor,...) to find new catering stores or to find things to do in the city respectively. Respondents who indicated to never use such websites or apps were not asked any further questions about these recommender systems and were forwarded to the next questionnaire section. All other respondents, in contrast, got follow-up questions about serendipity, satisfaction and conversion in these recommender systems.

Unfortunately, within the existing literature, examples of measurement instruments for serendipity experiences are scarce. Additionally, the few existing operationalizations are also quite diverse. Some use several agree-disagree statements for measuring specific dimensions of serendipity like perceived recommendation diversity (e.g. Knijnenburg, Willemsen, Gantner, Soncu, & Newell, 2012). Others developed item sets for measuring experiences about serendipity affordances based on the work of Björneborn (e.g. McCay-Peet & Toms, 2011) or implemented a set of questions allowing test persons to compare different recommender systems (e.g. Ekstrand et al., 2014). Some investigated the potential of survey questions for measuring a various range of serendipity definitions (e.g. Kotkov et al., 2018) or instantaneous experiences of serendipity (e.g. Lutz et al., 2017). Because of space constraints, however, we based our work on the survey items used by Chen et al. (2019), who adopted a short single-item version of the ResQue evaluation framework for recommender systems (see Pu et al., 2011).

In order to measure relevance of recommended items respondents were asked how often they get recommendations that suit them well (see the Appendix). For measuring novelty, they were asked how often they get recommendations they didn't know yet. For (lack of) diversity, they were asked how often they get the same kind of recommendations. For experienced serendipity, respondents were asked how often they find themselves pleasantly surprised by the recommendations in these systems. User satisfaction was measured by a question about how satisfied respondents generally are with the recommendations they usually get. User conversion was measured by a question about how often they actually follow the provided recommendations in such recommendation systems. Respondents could provide answers to all these questions through 5-point Likert scales. The Likert scales were treated as continuous variables in all analyses below (see Table 1).

3.3 Analysis

Given that our theoretical model (Figure 1) assumes an indirect effect of relevance, novelty and diversity on recommender system conversion through experienced serendipity and user satisfaction, we used mediation analysis to model the data. Mediation analysis refers to the investigation of direct and indirect effects of a set of exogenous independent variables on the dependent variables through mediator variables (MacKinnon, 2008).

Within the existing literature, mediation analysis usually starts from an investigation of the total effect of the independent variables on the dependent variables (see, for example, Baron & Kenny, 1986). Subsequently, the effect of the independent variables on the dependent variables is measured again but controlling for the mediator variables. This allows for discriminating between the direct and indirect effect of the independent variables on the dependent variables. Nonetheless, it is advised to model all effects simulta-

Table 1. The questionnaire included questions about serendipity, satisfaction and conversion of recommender systems in urban environments. The bivariate correlations between the serendipity antecedents, satisfaction and conversion are moderate to high, except for diversity.

			Correlations				
	Mean	Std. Dev.	Relevance	Novelty	Diversity	Serendipity	Satisfaction
Catering group							
Relevance	3.25	.81					
Novelty	3.40	.83	.41 [<.001]				
Diversity	3.35	.80	.05 [.140]	.09 [.009]			
Serendipity	3.03	.76	.58 [<.001]	.44 [<.001]	.00 [.900]		
Satisfaction	3.66	.71	.59 [<.001]	.41 [<.001]	.01 [.794]	.46 [<.001]	
Conversion	2.92	.74	.50 [<.001]	.37 [<.001]	.19 [<.001]	.46 [<.001]	.35 [<.001]
Activities group							
Relevance	3.33	.80					
Novelty	3.16	.78	.40 [<.001]				
Diversity	3.54	.78	-.07 [.047]	-.10 [.007]			
Serendipity	2.82	.79	.54 [<.001]	.48 [<.001]	-.17 [<.001]		
Satisfaction	3.74	.71	.64 [<.001]	.37 [<.001]	-.09 [.012]	.46 [<.001]	
Conversion	2.62	.83	.49 [<.001]	.40 [<.001]	-.06 [.066]	.57 [<.001]	.37 [<.001]

p-values between square brackets.

All responses were collected on 5-point Likert-scales. Question wording and response scales can be found in the appendix.

neously in one single analysis model (Rucker, Preacher, Tormala, & Petty, 2011). For that reason, we conducted a path analysis were paths were defined along the theoretical model in Figure 1 additional to direct effects of the serendipity antecedents relevance, novelty and diversity on the dependent variables user satisfaction and conversion. The model was estimated in `R` using the `lavaan` package (Rosseel, 2012). Multi-group estimation was used to distinguish between the Catering scenario and the Activities scenario. Parameter estimates were obtained by maximum likelihood estimation with robust Huber-White standard errors to avoid problems of non-normality.

Note that we do not correct *p*-values for multiple testing since this research is exploratory. The interpretation of results are mainly based on estimated effect sizes rather than *p*-values. After all, *p*-values are bad measures of effect sizes or the importance of results (Wasserstein & Lazar, 2016; Greenland et al., 2016; Amrhein, Greenland, & McShane, 2019). Moreover, adjustments for multiple testing increase type II errors for non-null associations and are calculated based on arbitrarily chosen numbers of tests (Rothman, 1990; Feise, 2002).

4 Results

4.1 Descriptive analyses

On average, the respondents provided similar answers to all survey items (see Table 1). Indeed, they were mostly undecided about whether recommender items are usually relevant, novel and diverse, whether they experience serendipity, are satisfied by the recommendations and adopt to the recommendations, because the observed means of all items are close to the center of the scale, i.e. 3. Nonetheless, except for conversion, the responses were slightly skewed to the positive side of the response scales in both the Catering as well as the Activities respondent groups. Additionally, the spread of the responses was more or less similar across all items in both experimental groups, as shown by the observed standard deviations.

When considering the bivariate correlations between the six items in both contexts, the data firstly reveal relatively strong relations between relevance, novelty and experienced serendipity. Diversity, in contrast, is very weakly correlated with experienced serendipity as well as with both other antecedents. This suggests that, in contrast to relevance and novelty, diversity in recommendations does not cause users to experience serendipity.

Regarding satisfaction with the outcomes of the recommender systems, the data yield the highest correlations for relevance. The correlations with novelty and experienced serendipity are also moderately high. Also here, there doesn't seem to be much correlation with diversity. This suggests that people prefer to obtain recommendations that are, above all, relevant, even though novelty may also increase satisfaction. Further analysis will reveal whether these effects can be explained by experienced serendipity. The results, however, do not suggest that users seek much diversity in their recommendations in order to be satisfied.

When considering user conversion, lastly, the data show slightly different patterns. Here, relevance doesn't seem to correlate much higher than novelty, even though all correlations are still moderately high. Also here, diversity doesn't seem to affect conversion at all in the Activities scenario, but it does show a moderate correlation with user conversion in the Catering scenario.

It should be noted that the correlation patterns are very similar between the two experimental groups, i.e. in the context of recommender systems for catering stores compared to recommender systems for urban activities. This suggests that the context of recommender system usage is of minor importance within these two scenarios.

Before the models were fit, we also tested for multicollinearity among the different items, because this may result in unreliable parameter estimates (Farrar & Glauber, 1967). Nonetheless, as none of the correlation coefficients exceeds .80, there is a low risk of multicollinearity.

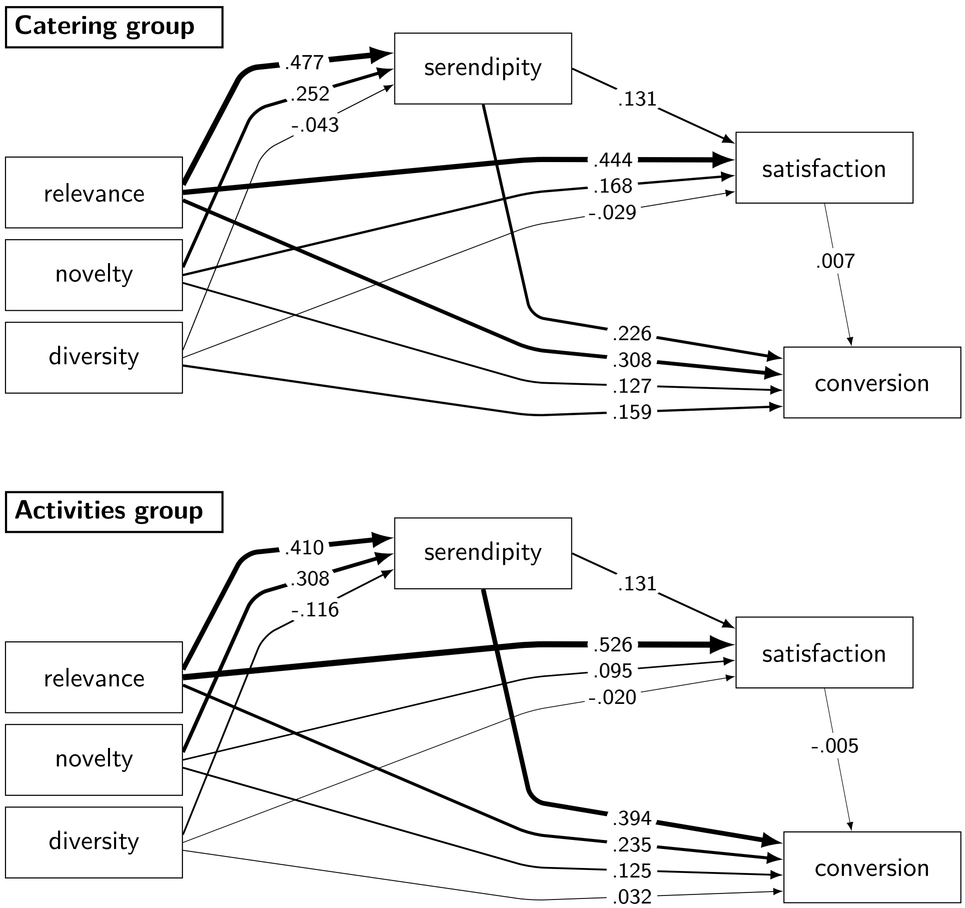
4.2 Mediation analysis

Considering the results of the path analysis (see Table 2 and Figure 2), the findings from the correlation matrix are confirmed. The results suggest that experienced serendipity largely depends on the relevance and the novelty of the recommendations, even after controlling for the other antecedents. Diversity in the recommendations, in contrast, doesn't seem to affect experienced serendipity. In the Activities group, the estimated effect of diversity on experienced serendipity was even slightly negative and statistically quite significant. All combined, the antecedents explain almost 40% of the variance in experienced serendipity in both the Catering and the Activities group, which is moderately high in social sciences.

Table 2. Relevance, novelty and serendipity don't seem to have an indirect effect on recommender system conversion through satisfaction, but they do have a direct effect.

	Stand. Effects on Serendipity	Standardized Effects on Satisfaction			Standardized Effects on Conversion		
		Direct	Indirect	Total	Direct	Indirect	Total
Catering group:							
Relevance	.477 [<.001]	.444 [<.001]	.063 [.024]	.507 [<.001]	.308 [<.001]	.003 [.914]	.311 [<.001]
Novelty	.252 [<.001]	.168 [.004]	.033 [.068]	.201 [<.001]	.127 [.019]	.001 [.913]	.128 [.037]
Diversity	-.043 [.248]	-.029 [.491]	-.006 [.233]	-.035 [.407]	.159 [<.001]	-.000 [.909]	.159 [<.001]
Serendipity		.131 [.030]			.226 [<.001]	.001 [.914]	.227 [<.001]
Satisfaction					.007 [.914]		
R^2	.388	.390			.339		
Activities group:							
Relevance	.410 [<.001]	.526 [<.001]	.054 [.002]	.580 [<.001]	.235 [<.001]	-.003 [.917]	.232 [<.001]
Novelty	.308 [<.001]	.095 [.007]	.040 [.002]	.135 [<.001]	.125 [.001]	-.001 [.917]	.124 [.001]
Diversity	-.116 [.001]	-.020 [.541]	-.015 [.014]	-.035 [.270]	.032 [.342]	.000 [.915]	.032 [.333]
Serendipity		.131 [.001]			.394 [<.001]	-.001 [.917]	.393 [<.001]
Satisfaction					-.005 [.917]		
R^2	.390	.433			.388		

p-values between square brackets.



Standardized coefficients are shown.

Figure 2. Our data suggest an effect of relevance, novelty and serendipity on recommender system satisfaction and conversion. Satisfaction doesn't seem to have an effect on conversion.

When considering the standardized effects on user satisfaction, relevance clearly shows the largest effect (.444 and .526 for the Catering and Activities group respectively). The more recommendations are relevant to the user, the more satisfied this user will be. The effect of novelty of the recommendations, in turn, is three to four times smaller (.168 and .095 respectively), although these effects are still highly significant. Similar to the effects on experienced serendipity, the effect of diversity on user satisfaction is almost non-existent in both experimental groups.

More surprisingly, however, the data do not confirm that the effect of relevance and novelty on user satisfaction can be explained by experienced serendipity. In both experimental groups, the indirect effects of relevance and novelty on satisfaction through experienced serendipity are very small. Put differently, our results do not confirm that higher user satisfaction due to high relevance and novelty in recommendations can be explained by positive experiences of serendipity. Also, after controlling for relevance, novelty and diversity, the effect of experienced serendipity on user satisfaction is much smaller compared to the bivariate correlations (see Table 1). This suggest that observed relations between experienced serendipity and user satisfaction are spurious correlations because they are both commonly caused by the relevance and novelty of recommendations.

Looking at the effects on user conversion, also here, the total effects of relevance and novelty remained fairly large and statistically very significant in both the Catering and Activities scenarios, even after controlling for the other variables including experienced serendipity and satisfaction. This firstly suggests that the effects of relevance and novelty on conversion can neither be explained by better experiences of serendipity nor by higher satisfaction with the recommended items.

Again, relevance of recommended items seems to have a larger effect on user conversion compared to the novelty of recommended items in both the Catering and Activities group (.308 versus .127 and .235 versus .125 respectively). Surprisingly, within the Catering group, the results also yielded a moderate effect of .159 for diversity. The more diverse recommended restaurants and bars are, the more the respondents state to follow up on these recommendations. Within the Activities group, in contrast, no such effect was found, similar to previous findings for diversity.

In contrast to the effects on user satisfaction, the results also yielded fairly large effects of experienced serendipity on user conversion. In the Activities scenario, the effect of experienced serendipity was even about twice as large as any other effect (.393 more specifically). These results imply that recommendations pleasantly surprising users increase the chance of users following up on these recommendations, and this can only marginally be explained by the relevance and the novelty of the recommendations.

The differences between the effects on satisfaction and user conversion seem to be explainable by a surprising lack of correlation between satisfaction and user conversion. Indeed, after controlling for relevance, novelty, diversity and experienced serendipity, the effect of satisfaction on user conversion completely disappears (that is .007 and -.005 respectively). Given the correlation between both constructs found in the bivariate analyses, these results suggest that satisfaction only relates to user conversion because both concepts are determined by relevance, novelty and experienced serendipity. In contrasts to our expectations, however, satisfaction does not seem to have an important influence on (self-reported) user conversion.

Last, it should be noted that no large differences were found between the Catering and Activities scenarios, except for the few effects already mentioned above. The purpose of the urban recommender system does not seem to have an important influence on the processes behind experienced serendipity in our examples.

5 Discussion

5.1 User evaluations of serendipity in urban recommender systems

This paper set out to study how users experience serendipity in urban recommender systems and which characteristics of the recommendations (novelty, relevance and diversity) contribute to this. Previous work had already studied experienced serendipity in recommender systems in various domains and identified some antecedents (e.g. Chen et al., 2019; Lutz et al., 2017), but few did this by a large-scale evaluation in the particular urban context. As a result, this paper presented a first exploration of the topic but also opened avenues for further research, which aligns with the relatively novelty of the study of serendipity in urban recommender systems.

Overall, our findings about the antecedents contributing to experienced serendipity in urban recommender systems are in line with findings from research in other application domains. The results of our research provide a clear affirmative answer to our first research question (*RQ1*): within urban recommender systems for catering stores and activities, experienced serendipity does primarily depend on the relevance of the recommended items and secondarily on the novelty of these items. Increases in diversity among the recommended items, however, do not seem to affect experienced serendipity. This is in line with other studies that also reported smaller effects for diversity (e.g. Chen et al., 2019).

However, when considering our second (*RQ2*) and third research question (*RQ3*) related to respectively user satisfaction and conversion, our results provided more complex conclusions. Firstly, the results of our survey did confirm that relevance and novelty positively affect user satisfaction with recommended items as well as the chance that users follow up on these recommendations, i.e. user conversion. For diversity, in contrast, such an effect was again completely absent. People thus get most satisfaction from recommendations when the recommendations are relevant and novel to them, while they barely care about any diversity in these recommendations.

Secondly, however, the effects of relevance and novelty on satisfaction and conversion can barely be explained by increases in experienced serendipity. Moreover, next to relevance and novelty, experienced serendipity also seems to have a separate effect on satisfaction and user conversion. These results might suggest that experienced serendipity should be considered as a construct that acts next to the experienced relevance and novelty of items in recommender systems. Future research may focus on the distinction and relations between all these concepts.

In addition, our results also revealed that user satisfaction and user conversion are only spuriously related because they commonly depend on factors like relevance, novelty, diversity and experienced serendipity of recommended items. After controlling for these factors, satisfaction and conversion surprisingly don't seem to be related at all. This unlikely lack of a direct relationship was, however, also found in previous studies (e.g. Lutz et al., 2017).

Finally, for the contextual differentiation under study in the fourth research question (*RQ4*), the results did not show large differences between recommender systems for catering stores and for urban activities. Put differently, the context of the recommendations does not seem to have an impact on the way serendipity related concepts determine user satisfaction and conversion. One notable exception was the effect of diversity on user conversion, which was moderately large in the Catering scenario and non-existent in the Activities scenario. It should be noted here, however, that the difference between recommender systems for catering stores and urban activities might not sufficiently reflect the diversity in urban recommender systems. Future research might thus elaborate on this topic.

5.2 Limitations

We are aware that our research may have some methodological limitations. The first limitation is the use of a retrospective survey about respondents' past experiences and behavior. Retrospectively asking people for their experiences may introduce measurement variance and bias because reported experiences may not correspond with true experiences at the moment the recommender systems were used due to recall bias. As already discussed above, a solution to this problem can be found in experimental studies. Nonetheless, such studies, while optimizing internal quality, may suffer from lower external quality because of the artificial settings they create. As a result, in order to get full insights in the topic of serendipity in urban recommender systems both experimental as well as observational studies are required in order to triangulate findings. For that reason, we believe our study provides a contribution to the field.

Another alternative to solve the problem of recall bias is the use of experience sampling, in which users are immediately asked some questions in a pop-up directly after using a recommender system. Experience sampling, however, is very difficult to implement as it requires adapting the recommender system user interface. Additionally, it may also annoy users leading to dropouts and it put some serious restrictions on the number of questions that can be asked.

The second limitation is that no causal relationships between the different concepts, as implied by the theoretical model (see Figure 1), could be proven by the study design. Such causal claims can only be investigated by true experimental designs in which system design characteristics related to relevance, novelty and diversity are explicitly manipulated by the study experimenter. Unfortunately, such experiments are still scarce in the existing literature. Moreover, such experiments also don't allow to make causal claims about the relationships between the true experiences of relevance, novelty, diversity and serendipity.

The third limitation is the fact that all concepts were measured by single questions instead of item batteries. Although this approach was also adopted in related work Chen et al. (2019) this may lead to measurement bias. Also, order effects may have played a role: the question about user satisfaction was asked as the first question while the conversion question as the last one. This might also explain the lack of correlation between satisfaction and conversion. Future research may focus on such order effects, for example, by randomizing the question order in order to neutralize such ordering effects. Further, question wording may also lead to measurement bias and variance due to interpretation differences among respondents. For example, the question measuring diversity may also be interpreted as how often the service updates recommendations for users. Likewise, novelty can be interpreted in different ways (Kotkov, Wang, & Veijalainen, 2016; Iaquinta et al., 2008, see) and it remains unclear to what extend the interpretation overlaps between the questions measuring novelty in particular and experienced serendipity in general. Unfortunately, research on proper question development about serendipity is still scarce and very heterogeneous in the current literature and may thus form an interesting topic for future research. Such research will not only entail questionnaire development and validation but will also require more in-depth research on the meaning of serendipity in urban environments (and urban recommender systems in particular).

Further, in order to circumvent an artificial research environment with a limited focus on certain antecedents of serendipity, we asked our respondents to think of the recommender systems they would actually use in their general daily life to find catering stores or activities within their cities instead of creating lab experiments to test for particular design differences in recommender engines. However, as a result of this strategy, we also don't know about which recommender systems our respondents thought about while completing the survey and to what extent their momentary reactions represent their overall opinions adequately.

Finally, our research strategy also depends on the current status of existing urban recommender systems. This means that our research merely involves an evaluation by citizens of these current systems rather than an investigation of how citizens actually want such systems to be. For example, it might well be the case that today's urban recommender systems lack a sufficient amount of diversity in their recommendations, which make questions about such diversity much more abstract to respondents.

A final remark should be made about the particular timing of our survey, which was during the early days of the global pandemic (March-April 2020). In Belgium, citizens were since mid-March restricted in their movements and so-called non-essential shops (including bars and restaurants) had to close. Despite this situation, we believe that it did not significantly impact our findings since the survey retrospectively questioned citizens about their past experiences and behavior. Since we asked these questions just at the beginning of the pandemic, citizens could still recall recent experiences. Nevertheless, if the timing were to have an impact, it could potentially explain the weak correlation between satisfaction and conversion. However, further work here is needed.

6 Conclusion

This work aimed to contribute to the existing work on serendipity by providing a first insight into user evaluations of serendipity in urban recommender systems. By means of a survey among 1641 citizens in Flanders and Brussels (Belgium) we collected data on their previous experiences with using recommender systems in urban contexts. More specifically, we explored to what extent characteristics of the recommended items (i.e. their relevance, novelty and diversity) led to experiences of serendipity and how this relates to user satisfaction and conversion.

Our findings showed that users' experiences of serendipity in urban recommender systems align with findings in other application domains in the sense that there is a strong relation between relevance, novelty and experienced serendipity. Moreover, serendipitous recommendations are found to increase the chance of users following up on these recommendations. A noteworthy finding is the fact that diversity is only weakly correlated with experienced serendipity, similar to findings in other work. We believe this result has to be interpreted carefully as it is exactly the assumed lack of diversity that spurs research into serendipitous recommendations. In other words, the lack of diversity in today's recommender systems might possibly explain this weak relationship.

By elaborating on the limitations of our study, we underlined the difficulty of collecting data on user evaluations of serendipity in urban recommender systems. We therefore call for further research that studies this subject in more depth, taking into account the previously suggested paths for further work. Such research will contribute to the current challenges that come along with the increasing implementation of technologies in our urban environments, and how this affects serendipity in the city.

References

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*, 305–307.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173–1182.

Björneborn, L. (2017). Three key affordances for serendipity. *Journal of Documentation*.

Chen, L., Yang, Y., Wang, N., Yang, K., & Yuan, Q. (2019). How serendipity improves user satisfaction with recommendations? a large-scale user evaluation. In *The world wide web conference* (pp. 240–250).

De Gemmis, M., Lops, P., Semeraro, G., & Musto, C. (2015). An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, *51*(5), 695–717.

Delva, H., Smets, A., Colpaert, P., Ballon, P., & Verborgh, R. (2020, June). Using regions of interest for personalized route planning. In *Proceedings of the 2nd international workshop on semantics for transport.*

de Waal, M. (2013). *De stad als interface: Hoe nieuwe media de stad veranderen*. Rotterdam: Nai010.

Ekstrand, M. D., Harper, F. M., Willemsen, M. C., & Konstan, J. A. (2014). User perception of differences in recommender algorithms. In *Proceedings of the 8th acm conference on recommender systems* (pp. 161–168).

Erdelez, S. (2004). Investigation of information encountering in the controlled research environment. *Information processing & management*, *40*(6), 1013–1025.

Erdelez, S., & Makri, S. (2020). Information encountering re-encountered. *Journal of Documentation.*

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, *49*(1), 92–107. Retrieved from `http://www.jstor.org/stable/1937887`

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC medical research methodology*, *2*(1), 1–4.

Foth, M. (2017). The software-sorted city: Big data & algorithms. In *Digital cities workshop.* ACM.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth acm conference on recommender systems* (pp. 257–260).

Ge, X., Daphalapurkar, A., Shimpi, M., Kohli, D., Pelechrinis, K., Chrysanthis, P. K., & Zeinalipour-Yazti, D. (2017). Data-driven serendipity navigation in urban places. In *2017 ieee 37th international conference on distributed computing systems (icdcs)* (pp. 2501–2504).

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European*

Journal of Epidemiology, *31*(4), 337–350.
Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008). Introducing serendipity in a content-based recommender system. In *2008 eighth international conference on hybrid intelligent systems* (pp. 168–173).
Jacobs, J. (1961). *The death and life of great american cities.* New york: Vintage.
Jannach, D., & Adomavicius, G. (2016). Recommendations with a purpose. In *Proceedings of the 10th acm conference on recommender systems* (pp. 7–10).
Jiang, T., Liu, F., & Chi, Y. (2015). Online information encountering: modeling the process and influencing factors. *Journal of Documentation.*
Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(1), 1–42.
Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, *22*(4), 441–504.
Kotkov, D., Konstan, J. A., Zhao, Q., & Veijalainen, J. (2018). Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd annual acm symposium on applied computing* (pp. 1341–1350).
Kotkov, D., Veijalainen, J., & Wang, S. (2016). Challenges of serendipity in recommender systems. In *Webist 2016: Proceedings of the 12th international conference on web information systems and technologies.* (Vol. 2).
Kotkov, D., Wang, S., & Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*, *111*, 180–192.
Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems–a survey. *Knowledge-Based Systems*, *123*, 154–162.
Li, P., & Tuzhilin, A. (2019, September). Latent modeling of unexpectedness for recommendations. In *Acm recsys (late-breaking results)* (pp. 36–40). Copenhagen, Denmark.
Lutz, C., Pieter Hoffmann, C., & Meckel, M. (2017). Online serendipity: A contextual differentiation of antecedents and outcomes. *Journal of the Association for Information Science and Technology*, *68*(7), 1698–1710.
MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* New York, NY: Taylor & Francis Group.

Makri, S., & Blandford, A. (2012). Coming across information serendipitously–part 1. *Journal of Documentation*.

Maksai, A., Garcin, F., & Faltings, B. (2015). Predicting online performance of news recommender systems through richer evaluation metrics. In *Proceedings of the 9th acm conference on recommender systems* (pp. 179–186).

Matt, C., Benlian, A., Hess, T., & Weiß, C. (2014). Escaping from the filter bubble? the effects of novelty and serendipity on users' evaluations of online recommendations. In *Icis2014 proceedings*.

McCay-Peet, L. (2014). *Investigating work-related serendipity, what influences it, and how it may be facilitated in digital environments* (Unpublished doctoral dissertation).

McCay-Peet, L., & Toms, E. (2011). Measuring the dimensions of serendipity in digital environments. *Information Research: An International Electronic Journal*, *16*(3), n3.

McQuire, S. (2017). *Geomedia: Networked cities and the future of public space*. John Wiley & Sons.

Olshannikova, E., Olsson, T., Huhtamäki, J., Paasovaara, S., & Kärkkäinen, H. (2020). From chance to serendipity: Knowledge workers' experiences of serendipitous social encounters. *Advances in Human-Computer Interaction*, *2020*.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin UK.

Paulos, E., & Goodman, E. (2004). The familiar stranger: anxiety, comfort, and play in public places. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 223–230).

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth acm conference on recommender systems* (pp. 157–164).

Quijano-Sánchez, L., Cantador, I., Cortés-Cediel, M. E., & Gil, O. (2020). Recommender systems for smart cities. *Information Systems*, 101545.

Reviglio, U. (2019). Towards a taxonomy for designing serendipity in personalized news feeds. In *Proceedings of the 10th international conference on conceptions of library and information science.* University of Borås.

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software, Articles*, *48*(2), 1–36. Retrieved from `https://www.jstatsoft.org/v048/i02` doi: 10.18637/jss.v048.i02

Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, 43–46.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, *5*(6), 359–371.

Said, A., Fields, B., Jain, B. J., & Albayrak, S. (2013). User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm. In *Proceedings of the 2013 conference on computer supported cooperative work* (pp. 1399–1408).

Shepard, M. (2011). *Sentient city: Ubiquitous computing, architecture, and the future of urban space.* The MIT press.

Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2019). How good your recommender system is? a survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, *10*(5), 813–831.

Smets, A., Montero, E., & Ballon, P. (2019). Does the bubble go beyond?: An exploration of the urban filter bubble. In *The 1st workshop on the impact of recommender systems with acm recsys 2019.* CEUR Workshop Proceedings.

Sun, X., Sharples, S., & Makri, S. (2011). A user-centred mobile diary study approach to understanding serendipity in information research. *Information Research*, *16*(3), 16–3.

Tacchini, E. (2012). *Serendipitous mentorship in music recommender systems* (Unpublished doctoral dissertation).

van der Graaf, S., & Ballon, P. (2019). Navigating platform urbanism. *Technological Forecasting and social change*, *142*, 364–372.

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157–178.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129-133.

Wood, P., & Landry, C. (2008). *The intercultural city: Planning for diversity advantage.* Routledge.

Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012). Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth acm international conference on web search and data mining* (pp. 13–22).

Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review. Journal on Internet Regulation*, *5*(1).

Appendix

Questions Catering group

1. Searching for a bar or restaurant on the internet. How do you experience this? How often do you use websites or apps (e.g. search robots like Google, TripAdvisor,...) to search for new restaurants or bars? Very often — Often — Sometimes — Rarely — Never
2. How satisfied are you with the recommended restaurants or bars on these websites or apps? Very satisfied — Satisfied — Neither satisfied, nor dissatisfied — Dissatisfied — Very dissatisfied
3. How often do you think such websites or apps recommend restaurants or bars that suit you well? Very often — Often — Sometimes — Rarely — Never
4. How often do you think such websites and apps recommend restaurants or bars you don't know yet? Very often — Often — Sometimes — Rarely — Never
5. How often do you think you get the same kind of restaurants or bars recommended on these websites and apps? Very often — Often — Sometimes — Rarely — Never
6. How often do you find yourself pleasantly surprised by the recommended restaurants or bars on these websites and apps? Very often — Often — Sometimes — Rarely — Never
7. How often do you actually go to the recommended restaurants or bars on these websites and apps? Very often — Often — Sometimes — Rarely — Never

Questions Activities group

1. Searching for activities on the internet. How do you experience this? How often do you use websites or apps (e.g. search robots such as Google, TripAdvisor, ...) to find out what to do in a city (e.g. when you are on vacation or planning a day trip)? Very often — Often — Sometimes — Rarely — Never
2. How satisfied are you with the recommended restaurants or bars on these websites or apps? Very satisfied — Satisfied — Neither satisfied, nor dissatisfied — Dissatisfied — Very dissatisfied

3. How often do you think such websites or apps recommend restaurants or bars that suit you well? Very often — Often — Sometimes — Rarely — Never
4. How often do you think such websites and apps recommend activities you don't know yet? Very often — Often — Sometimes — Rarely — Never
5. How often do you think you get the same kind of activities recommended on these websites and apps? Very often — Often — Sometimes — Rarely — Never
6. How often do you find yourself pleasantly surprised by the recommended activities on these websites and apps? Very often — Often — Sometimes — Rarely — Never
7. How often do you actually go to or actually participate in the recommended activities on these websites and apps? Very often — Often — Sometimes — Rarely — Never