Aalto University
School of Business

The Future of A/B Testing in Social Network Advertising

Bachelor's Thesis
Jenna Nyländen
4.1.2018
Information Systems Management

Approved in the Department of Information and Service Economy xx.xx.20xx and awarded the grade

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of bachelor's thesis

Author Jenna Nyländen

Title of thesis The Future of A/B Testing in Social Network Advertising

Degree Bachelor of Science in Economics and Business Administration

Degree programme Information and Service Management

Thesis advisor(s) Niina Mallat, Jani Merikivi

Year of approval 2018 **Number of pages** 27 **Language** English

Abstract

This research addresses the future of A/B testing in social network advertising. A/B test is a well-studied comparison problem with two different samples with the goal of testing the treatment effect of old and new variations. In recent years, through the rise of the internet, A/B testing in social networks has gained sharpened focus and is commonly used in social network advertising. Due to the market-driven strategy the companies should today aim for, the development of A/B testing in social network advertising can help in gathering useful insights of consumer preferences and attitudes. A/B testing has been perceived as cheap, simple and reliable way of optimizing advertisement and mining data from site users. However, as currently performed A/B testing has criticized as manual and time-consuming activity that requires complex set of statistical and engineering skills. This study focuses on overcoming these problems through automation and machine learning algorithms. Besides, the importance of shifting organizational focus on optimal usage of data-driven decision making through A/B testing, and user attitudes towards social network advertising and their ad-clicking behaviour are addressed.

Keywords A/B testing, social network site, social network advertising, market-driven strategy, data-driven decision making, consumer attitudes, ad-clicking behaviour, automation, machine learning algorithms

Table of Contents

1 Introduction

A/B testing, also known as splitting testing, bucket testing or controlled experiment, has been widely used for evaluating new feature in the data-driven decision-making processes of online websites, especially those in social networks sites (SNS). The goal of A/B testing is to estimate the difference between the treatment effects of different variations. It is a well-studied comparison problem with two samples. In the simplest form of A/B testing, the aim is to evaluate the effect of a new feature within the site users compared to the old version of it. The new feature is exposed to a small randomly selected fraction of the user population and its effect is then measured. Most of the A/B tests concentrate on the aspects that are visible to the user, front-end of the application, including layouts, fonts, colours, etc. (Backstrom & Kleinberg, 2011; Cubero et al. 2016; Geng et al. 2016).

The roots of A/B testing lie in the 1700s. Back then a British ship's captain realized sailors staying healthier and avoiding scurvy when sailing to Mediterranean countries where fruit was easily accessible. Based on this notation he then gave half of his crew limes (treatment group) and the rest of the crew continued with their regular diet (control group). Experiment was successful and the captain figured that the British sailors should consume citrus fruits regularly. This example can be seen as an early state of A/B testing. The treatment group eating limes is the option A and the control group is the option B in the comparison problem. Many years later, in the 2000s, the first A/B tests were run on the web. (Henne et al. 2008).

Since the rise of the internet, A/B tests have expanded everywhere (Adegeest et al. 2017). In recent years, A/B testing in social networks has gained sharpened focus and is commonly used in social network advertising (SNA) in companies such as Facebook, LinkedIn and Twitter (Geng et al. 2016). For network A/B test, in particular, Glass et al. (2016) have widened the theory of traditional A/B testing by suggesting that the responses of users in an A/B test are affected by each other's choices when acting in the same network with each other. Hence, they introduce the concept of network effect which results from the correlation of a user's and her neighbours' behaviour.

It is argued that the method of A/B testing is simple, cheap and reliable way to test the consumer needs, attitudes, navigation profiles, and preferences (Adegeest et al. 2017). A/B testing not only offers a test pattern for optimizing website advertisement but also mines large amounts of data about the behaviour of consumers and users (Henne et al. 2008). Hence, through A/B testing, companies are able to gather thorough insights about consumer interests and the factors influencing the outcome of those. As one aspect to the benefits of A/B testing stands its support for the company maintaining and developing an experimentation culture. It has been argued that an experimentation culture is a certain way to get to the bottom of a decision without relying on anyone's opinion as more of ideas will be driven further when presented in the form of tests and workers will be more highly motivated because they get to see ideas live in the real world (Jenkins, 2014). This is the managerial approach to the benefits of utilization of A/B testing in a company working on any field.

1.1 Latest development in the field and the key concepts

Adegeest et al. (2017) criticize traditional A/B testing due to its tendency to presume that the companies conducting tests only have enough resources and capacity to keep one of the two test variants at once. If more capacity for the A/B tests is available in forms of more advanced data science techniques, it is possible to gain more profit for the company from the results. To answer this limitation, Adegeest et al. developed the concept of A&B testing in which the effect of users preferring the "loser-option" is not ignored and thus valuable conversions are not left out in the decision-making process. This occurs through the automation of website personalization based on the findings of Exceptional Model Mining (EMM), an advanced data science technique that identifies subgroups behaving differently from the overall population.

Scott W. H. Young (2014), in turn, argues that when employed in isolation from other related functions, A/B testing can lead into incomplete conclusions. Thus, it is necessary integrate A/B testing into the entire market research program of the company. Moreover, recent study from Cubero et al. (2016) has shown that the way A/B testing is used only for testing usability aspects such as creating variations of the graphical user interface (different layouts, fonts, colours, etc.) is limiting. They introduce a tool for testing the aspects related to many different business processes so that A/B testing could be exploited to support and improve a larger set of systems with continuous verification and validation infrastructure that allows to experiment with the involvement of end users.

As the most recent development and discussion in the field of A/B testing shows, the rapid increase of advanced data science techniques, the synchronization of different processes within and between the business activities, and the concrete need of the iterative involvement of end user in the corporate decision making processes offer both opportunities and challenges for traditional A/B testing. However, in its current manner, traditional A/B testing has been seen as a time consuming, error prone and costly manual activity when performed in a large scale. Even when successfully executed in the whole industry, A/B testing is considered by the majority as a complex, manual and costly activity rather than a sophisticated software engineering process, according to Margara and Tamburelli (2014). They demonstrate the practical feasibility of automated A/B testing through a set of synthetic experiments. However, it is still not clear, whether the companies are moving towards the given automatic A/B testing models in their daily practices or not.

From different terms to describe the method of evaluating new feature in the data-driven decision-making processes of online websites, I choose to use 'A/B testing' as it is the one that is easily associated with the variation testing in the advertising field and is used widely in current discussion about the subject. Moreover, I abbreviate the terms 'social network site' and 'social network advertising' as SNS and SNA, respectively.

1.2 Motivation for work

The importance of studying the topic of A/B testing in SNA context in the near future lies in the market-driven strategy companies should today aim for. A market-driven strategy supports the company's understanding of its own market and customer-base, and the planning of all business processes based on these insights. Customer satisfaction and customer value are increased which again affects positively company's return on investment (ROI) and profitability. (Leventhal, 2017).

Furthermore, the growing possibilities in the field of advertising due to the trends of SNSs and big data underline another aspect for studying the method of A/B testing further. When a (potential) customer connects with the brand through an SNA it is possible to make the advertising decisions based on the data collected through these connections and when these touch points occur via internet, social media, and mobile devices, data can be saved over time for huge amounts of users (Li & Malthouse, 2017).

As the volume of this data grows rapidly, focus needs to be shifted into how to treat the A/B test data and make decisions based on it. As third aspect, monetizing big data has been discussed a lot recently. All of the points covered underline the importance of continuing the research of A/B testing in the context of SNA. The increasing focus on marketing and advertising optimization and the need for collecting data about consumer needs and preferences offer the market pull, and the new technological development offers the technology push for the innovations regarding A/B testing.

Based on the recent development, the concept of A/B testing as a test model and its improvements regarding practical factors such as choosing what to test, defining different audiences effectively, choosing the correct framework and interpreting the results have been widely discussed. Besides, researchers such as Deng et al. (2014) have introduced a wide set of "rule of thumbs" for online network A/B testing. However, the question about the effectiveness of A/B testing is no more only about whether the company understands the importance of A/B testing as a whole or is able to design and implement it but the more accurate dimensions lie in the new aspects that the most recent technological development enables. The limitations of previous study lie in the companies' ability to analyse and utilize the A/B test data, consumers' attitudes and ad-clicking behaviour addressing A/B tests, and the concepts of automation of manual work and machine learning algorithms. These factors have little to do with how the test is conducted in practice but with new managerial and technological possibilities for improving A/B testing in the future. Thus, in this paper the focus is shifted from the A/B testing itself to the newest technological turning points, managerial approaches towards the data and consumer preferences, and how those will transform how companies perform their advertising optimization and the data mining related to it in the future.

1.3 Research objectives and research questions

The objective of this research is to study the possible future scenarios for A/B testing in the field of SNA. The aim is to examine the method of A/B testing with respect to the directions of the technological development according to the literature and how those will most likely affect the advertising optimization and data mining in SNSs in the near future.

Specific research questions:

RQ1: How will A/B testing in the context of social network advertising most likely develop in the near future?

RQ2: What are the implications in the recent technological development that affect firms' advertising optimization and related data mining methods the most?

1.4 Scope of the research

In this research, I will focus on A/B testing in the context of social network advertising. This excludes all the insights and implications of A/B testing in any other fields of business. The concept of social network in this research includes different online SNSs, such as Facebook, Instagram, Twitter and LinkedIn. By using the term social media, I exclude all the online marketing functions that don't occur in the shared social engagement platforms, social network sites. Thus, for example, company websites and optimization of those is left out. The focus is shifted to the elements that are relevant in the network advertising context, including visual creativity aspect (i.e. the ad picture), ad text and the links related to the ad. Also, no marketing activities that occur offline will be covered. Furthermore, all the other marketing functions besides advertising are excluded in this research. Thus, the focus can be targeted to the modern advertisement optimization and data mining through SNA. In fact, these two benefits that can be gained by using the method of A/B testing are both focused on in the study.

The research objective is to study how A/B testing will most likely develop in the future and how current technology transformation will answer today's issues regarding advertising optimization and data mining in online social networks through A/B testing. Thus, I will go through different technology and business trends in a large scope, including factors such as automation. Roles of both company employees working with the tests and SNS users will be taken into consideration in practical and attitude levels.

1.5 Structure of the research

The rest of the thesis is structured as follows. Chapter two presents the methodological approach for this study. I will explain the data used in this research and the method of searching it. I will specify the different databases and sources, explain how I searched the data in practice and introduce my article inclusion criteria. Besides, I will cover the ethical standards concerned in my thesis.

In chapter three I will present the results of my work. The suggestions based on the previous literature are described. The chapter has five second-level subsections in each of which I approach SNS A/B testing and the development of it from different viewpoints that I found most important during my research. This also includes the insights collected from the case company I conducted a question patter to.

Finally, chapter four covers the discussion and final conclusions of my work. I will summarize and conclude the key findings. I will also focus on what the results indicate in both research and practice when considering companies and other actors operating in the field and topic area, and make recommendations based on my findings. Lastly, I will end this paper with the limitations of this study and by giving implications for future research in the field.

2 Methodology

The method of my thesis is a literature review. Thus, my data is the previous literature in the field. Besides, I used insights from a case example company. The order and structure to how my research has been undertaken is explained in this section. In general, I collected published articles related to the topic of A/B testing in social network advertising and analysed what can be learned through considering those collectively. I cite Denney's and Tewksbury's research (2003) 'A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research' throughout the chapter.

I first explain where I searched for the articles that are included in my work. Second, I answer how I searched my data in practice. Third, I go through the article inclusion criteria that were used when selecting the articles that are included in my work and the ethical issues that assign the collection of the data. Finally, I consider the data addressing the case example company I used in my work in order to gain more practical insights about the topic.

2.1 Databases used in my work

The main place to search for the articles I used in my work were all the Aalto Learning Centre's resources in the Aalto-Finna portal. Denney and Tewksbury argue the online databases at universities' library websites being the main way of finding sources for a literature review. Besides the Aalto-Finna portal, the Scopus and Google Scholar databases offered relevant data for my work in some of the cases. From these collections I included mainly academic journal articles, as those are the most appropriate source of information (Denney and Tewksbury, 2003). Besides, I used for example newspaper articles, magazine articles and government publications. Denney and Tewksbury argue these also being appropriate sources.

2.2 Keyword search and backward search

When searching for the articles I used keyword search. According to Denney and Tewksbury, this is the first step of the literature search process, however, it should not be the main part. I defined a set of search words I used collectively throughout the process. When searching previous literature about A/B testing itself as a methodology, I used the search word 'A/B testing' at most. Besides, I used its synonyms, such as 'controlled experiment', 'splitting testing' and 'bucket testing', as the usage of the term might have changed over time. Also, by changing the typology of the word 'A/B testing' (i.e. ab testing) in some cases lead to relevant results. Besides the articles concerning precisely A/B testing, I tried out many search words that are or might be related to the field and combinations of those, for example 'online advertising' and 'big data', 'advertising optimization' and 'ctr' or 'click-through rates', 'artificial intelligence' and 'advertising', etc.

As Denney and Tewksbury suggest, it is sometimes useful to get deeper into the citations, keywords, and authors of the previous sources to find new useful sources and insights through them. This approach is called backward search and is used in my thesis as well. In practice, as any specific key words, for instance 'ctr', drew my attention while reading any of the articles found during the keyword search, I would use it as the next search word. Same occurred with any highlighting authors or citations.

2.3 Criteria for the source inclusion

Denney and Tewksbury highlight the volume of the available literature being dependent on the topic. As the field of my study regards the modern world advertising optimization and data mining, I included the articles I used in my thesis mainly based on the criteria of their newness. In the means of collecting insights for the possible future development of A/B testing, very rare older articles can provide value in my thesis. In most cases I limited my search so that it offered only articles from the past three years or less (2015-2017). This criterion addresses especially the searches regarding the technological development – when trying to find implications of the future scenarios. However, when studying the A/B testing in general and other settled theories, I accepted also older articles. Denney and Tewksbury support this by arguing that it is important to include both classic and recent studies into the literature review. Nonetheless, the oldest article used is from the year 2003 and thus it can be said that all the data used is in the scope of past 15 years.

Both qualitative and quantitative research articles could be included and there were no geographical limitations. Also, the articles didn't necessarily have to be directly related to the topic of A/B testing as my research requires insights and views of the general technological and business development that might affect A/B testing somehow in the near future.

Ethical standards were also concerned in the thesis. The ethical issues regarding my literature review include factors such as information having to be obtained lawfully and reported accurately. I treat the work of existing researchers accurately and fairly which includes for example accurate citations and references.

2.4 Case company insights

In my thesis I compare the insights arising from the previous literature in the field to the practical side in forms of a case example company. The company is a middle sized Finnish listed company that operates globally and in the field of e-commerce. They use A/B testing as a relevant part of their marketing and advertising strategy.

I conducted a detailed question patter to the company's CDO. The questions were open, qualitative, questions including general aspects regarding the SNA channels, software used in for the process, variation objects (i.e. fonts, textual details or visual details), and the automation of the process. Besides, I included questions on how the tests are conducted in more detail including the sample size, amount of different test variants, the tracking of the results etc.

3 Results

In this chapter, the results of my research, the suggestions based on the previous literature, are described. I approach SNA A/B testing and the development of it from several different viewpoints, including both managerial and practical approaches. My findings are explained and structured as follows. First, I compare the insights of organizational capabilities on data and business insights, according to the existing literature, to the method of A/B testing. I conclude these problems in the level of employees', such as developers', designers' and analysts' regular daily work. Second, I move to the results that arise from the research on user attitudes towards SNA and their ad-clicking behaviour. Third, I continue with the challenges and possibilities regarding A/B testing in developing a personalized user experience and customized advertising. Fourth, I report my findings regarding the transformation of A/B testing methods in the future due to the most recent technological development. These include the possibilities and challenges regarding automation of manual work and machine learning algorithms. Finally, I will consider the result I gained from the case company and this way include and example from the real life practical side into the work.

3.1 Problems arising from practical daily work and organizational perceptions of data

When studying the future of A/B testing in SNA context, it is relevant to consider it as a method that a company of any size and any field could include in their advertising strategy and implement in practice. Here, the experimentation culture and attitudes towards measurement in general within the organization become accurate. In the history of social science, it is emphasized that simply practicing measurement as an absolute process does not necessarily provide useful and relevant insights the organization can use to improve its processes and performance (Lawler et al., 1985; Mohrman et al., 2011). Increased measurement does not guarantee insights that can be developed into practice. Modern software systems deal with continuously growing and changing user population that may add up to millions requests daily (Margara et al., 2014). It is easier than ever to experiment with expanded volume of both potential and accurate data available to organizations when conducting A/B tests. Thus, the careful definition of which parts of this data to include in the company decision-making processes and which to leave out underlines an important but also difficult part of the testing process. For instance, in the book ‘A/B testing: The Most Powerful Way to Turn Click into Customers’, Koomen and Siroker (2013) argue that improved conversions from A/B tested elements that are not considered carefully enough, might cost the company something more valuable, such as their brand. For instance, consider a company conducting A/B test in order to change the colouring of their recent ad campaign but it results in the user base preferring an option that is not recognizable as a part of the company’s branding and marketing strategy. Even if the “winner-option” would gain a larger click-through-rate, its revenue is taken from the company’s previous and current branding effort. Thus, too blind-eyed focus on A/B test data can lead to results that are not in line with the overall advertising and marketing strategy. Indeed, Fink and Levenson (2017) argue that it is much more important to challenge the data sources, their methods of measurement and quality, instead of focusing only on the general business importance of what is being tested. Also, the ways of this data being interpreted and defined correctly cannot be underestimated.

Besides the organizational perceptions of data, interpreting it, and making decisions based on it, there are difficulties regarding the more practical side of A/B testing work. Magara and Tamburelli (2014) illustrate the most critical of these tasks faced by the developers of A/B tests as follows:

1. Development, deployment and modification of multiple variants continuously and real time for the successful implementation and evaluation of these variants.
2. Defining a variant means the consideration of how many variants and which ones of them to change when creating a new variant.
3. In traditional A/B testing, there are two variants at once. However, it is possible to simultaneously deploy more than two variants which, in fact, constitutes the A&B test by Adegeest et al. (2017). Challenges occur when having to choose the amount of different variants and to modify it over time.
4. A/B testing works iteratively. Selecting the most potential variants in the beginning of each iteration round and prioritizing some variants over another might not be an easy task, as programs can be large and complicated.
5. Evaluation of variants requires mathematical and statistical skills that the workers developing the code itself might not have.
6. Knowing when to stop the tests and understanding when optimal solution has at least nearly been reached is crucial for using time and effort only for relevant changes in certain specific aspects.

These tasks demonstrate the complexity and challenges regarding the manual activity of A/B testing and the diversity of skills required from the employees to perform the tests effectively. I continue the results regarding the solutions to these problems in the section 3.4. that addresses the technical development of A/B testing such as the possibility to automate these manual tasks.

What it comes to the process of analysing and criticizing the data, Fink and Levenson see it as a function of the senior management. However, the previous literature of A/B testing offers an opposite insight. The concept of Highest Paid Person's Opinion (HiPPO) is widely considered in the research field of A/B testing and describes the strong managers with strong opinions but with the lack of data. No substantial learning and and return on investment (ROI) occur before the company employees and management listen to their customers, not to anyone's intuition or opinions (Henne et al. 2008). Thus, before the method of A/B testing and the decision-making processes based on it can be seen optimal for any sort of companies, the careful implementation of experimentation culture within the organization has to be executed. Recent EY surveys support this idea by arguing that 81% of executives say they believe that "data should be at the heart of all decision-making". On the other hand, Golsby-Smith and Martin (2017) argue that only having data is not any kind of proof that outcomes couldn't be different from the insights of this data. Moreover, that data is no more than an evidence and it's not always obvious what it is an evidence of. In fact, Golsby-Smith and Martin propose that the belief that all business decisions should be reached through scientific analysis decreases strategic options and innovation in the company. They argue that scientific methods are designed to understand natural phenomena that cannot be changed and thus, is not an effective way to evaluate things that do not yet exist. This, as well, is a perspective the CDO's and CMO's of the companies should take into account when conducting A/B testing in its current manner. In fact, in the case of SNA A/B testing, this perspective relates strongly to the effectiveness of prototyping, the creation of the variety of advertisement alternatives that set the starting point for the whole testing process. The following chapter introduces guidance to the problem of choosing the most potential advertising prototypes from early on according to previous literature on user attitudes and their ad-clicking behaviour.

3.2 Beliefs and concerns on user attitudes towards online social network advertising and their ad-clicking behaviour

There are recommendations arising from previous literature on SNA that can be applied in the creation of A/B testing prototypes and the whole process in general. When trying to figure out what kind of advertisement prototypes to include in the A/B testing processes, several factors regarding the user attitudes and reactions in general should be taken into consideration. Baek's and Morimoto's (2012) research highlights two important factors. First, "consumer concerns of advertising as intrusive and irritating affect their attitudes toward online advertising". Second, "User attitudes toward online advertising affect their ad-clicking behavior". Moreover, Lewin et al. (2011) argue that "For advertisers and the sites themselves, it is crucial that users accept advertising as a component of the SNS". If users' attitudes shift onto preferring to not viewing or seeing SNA at all, optimization of the ads through A/B testing will lose its purpose. On the other hand, through A/B testing the problems arising from these factors might be possible to reduce. This requires companies to consider the "loser options" of the running tests not only as unsuccessful prototypes but also as ads that are only targeted to the audiences avoiding SNA. In this way the existing understanding of A/B testing in SNA as a method of learning customer preferences efficiently can be deepened.

The current stage of the research implicates that user's attitudes and ad-clicking behaviour are not threatening SNA in the near future. Mir (2015) found that site users see SNA as beneficial for the economy which increases their positive attitudes toward it. On the other hand, the same study indicates that SNS users perceive advertising in such sites as misleading, supporting materialism and corrupting social values. However, these beliefs don't seem to have negative effects on SNS users' attitudes towards the advertising. Nevertheless, consumers vary from individual to individual in terms of factors that are not easy to observe, including their ad clicking behaviours which could be affected by their inherent purchase intention, exposure to marketing communication, or preference for one advertising format over another (Duan et al. 2014). Thus, it is important to include consumers' individual heterogeneity into the testing model. This, indeed, is the part where A/B testing makes relevant contribution in the creation of the valuable SNA prototypes.

3.3 The role of A/B testing when generating and improving personalized visitor experience and customized advertising

The interactions with the website can be instrumented in many different ways. Deng et al. (2014) suggest page views or clicks for this purpose. Also, there are several options for computing the key metrics. The most recommended ones include ad-clicking or click through-rates, sessions per user or revenue per user. Ad-clicking and click through rates indicate the direct response to the specific SNS advertisement for obtaining comprehensive product information (Cho, 2003). Also, the computations of these actions underlie an important recommendation process that creates a browsing path for the SNS users (Besbes et al. 2014). This particular research addresses the browsing path of online newspaper articles. It suggests that the website administration holds a database of feasible articles which includes information on for instance the topic classification, the publish date, or the click history. The available information is processed by several competing and complementary algorithms that analyse different aspects of it: the contextual connection between the host article and the candidates for recommendation; the reading behaviour and patterns associated with articles and readers; and additional information such as general track trends in the content network. These inputs are combined to generate a customized content recommendation. In the advertising optimization context, the same methodology is possible to apply. Similar information from advertisement in SNSs can be collected, processed and analysed, as the role of an article is replaced by an ad. Thus, recommendations of advertisements could be made to create a similar kind of path. By constructing recommendations for a future path, the waste of advertisements with low click through-rates could be reduced.

Due to Duan et al. 2014, there are sophisticated models that can analyse the conversion effects of website visits and advertisement. Through these models it is possible evaluate quite precisely the conversion effects and forecast the probability for purchases. This is due to the models accounting the entire clickstream history of individual users as a result of the accumulative effects of all previous click. This leads to scenarios where companies with careful data analysis are capable of identifying individual users and their ad-clicking paths which is the key to the creation of personalized visitor experience on SNSs and targeting advertisement individual by individual. However, this power of corporates leads to questions regarding SNS users' privacy and their concerns regarding their spreading data. Yet, the research on A/B testing has not concerned privacy issues that can arise from this specific area. Questions remain whether A/B testing in SNA can even collect such personal data that could cause privacy concerns in the future. The topic has to do with the level of preciseness of the data and the conclusions than can be made after large enough data sets are collected and analysed.

In the Wired Magazine, Brian Christian (2012) emphasizes an important insight: “Today, A/B is ubiquitous, and one of the strange consequences of that ubiquity is that the way we think about the web has become increasingly outdated. We talk about the Google homepage or the Amazon checkout screen, but it’s now more accurate to say that you visited a Google homepage, an Amazon checkout screen.” Just as the websites of these large companies can be personalized for each user separately, so could be made for the SNA. As the previous research indicates, A&B testing and Exceptional Model Mining (EMM) set the basis for this kind of unique personalization becoming possible. The A/B test data is no longer for choosing one specific option from several alternatives but utilizing an own “winner-option” for each user separately and deploying them one-by-one.

3.4 The possibilities of recent technological development

In this subchapter I go through the possible future implications of A/B testing from the perspective of recent technological development. I include automation of manual work and machine learning algorithms in this section as those arose as the most crucial ones when studying the limitations of A/B testing in its current manner and the possible future implications.

3.4.1 Automation of manual work in A/B testing

Despite A/B testing becoming ubiquitous, the technological side of it has not yet become straightforward. It requires complicated technological skills to gather insights of user actions and then rearrange the advertisement on the go based on these insights. Interpreting the results of A/B tests requires deep knowledge of statistics. Even the tools provided by third parties for this purpose, might require developers to code large sets for both A and B variants. This often results in nonprogrammers, such as marketing, editorial or design employees not being able run the tests without first outsourcing their tasks to engineers who then have to use their time for writing code for the several testing prototypes and variants. This, again, results in huge delays in seeing results as companies wait for this outsourced tasks to be finished and then for the ready products to go live. (Christian, 2012). Section 3.1. defined problems regarding the difficulties and issues in the usage of traditional A/B testing. These are emerging in a large part from a limited degree of automation in A/B testing which is the limitation this subchapter aims to answer. Moreover, I will cover the problem of manually calculating the statistical significance of A/B test and the options that automation offers for solving that part of the process.

To answer the problem of non-automated A/B testing, previous literature indicates that by formulating A/B testing as an optimization problem it is possible to implement automated search algorithms that work in real SNSs. This occurs through specifying advertisement features with a design-time declarative facility and a real-time framework that automatically and iteratively searches for possible concrete programs by generating, executing, and evaluating variants. In the step of specifying features, the developers write parametric program so that it covers all the relevant feature options and variants. This has to be done only once and then the model will automatically set them up on the site at the right time. Thus, huge amounts of work arising from developing and deploying potential variants is reduced. Through the step of selecting and evaluating variants, on the other hand, the model then tracks the actions on the site and makes decisions itself through continuous search. This decreases the amount of manual work in guiding and controlling the ongoing A/B tests. This method has been proved to work in practice, although not on real users. The model is able to shift towards an optimal solution and it always got closer to this solution with a small enough number of trials when tested in practice. (Magara and Tamburelli, 2014).

What it comes to the human work of calculating statistical significance of A/B testing, employees are not allowed to continuously track the results of ongoing tests and make decisions real time. This limitation has been seen as a step backwards in a world where technology should be able to conduct data analytics real time. Most A/B tests are conducted using the statistical theory of Null Hypothesis Statistical Testing (NHST), t-test or z-test. The results are interpreted by describing a specific significance level in the beginning of the test, and rejecting the null hypothesis if the p-value is smaller than the desired significance level. Checking test outcome while the test is running might change the outcome of the test itself as the calculated p-value fluctuates during the data is collected. (Deng et al. 2016).

This issue is called an optional stopping problem. However, recently automating A/B testing through Bayesian Hypothesis Testing has gained increasing interest as it is more suitable for real time decision making than NHST and overcomes the optional stopping problem. The main idea is that if the decision to stop depends on the outcome, whether a difference exists, or how large the difference is, the test result will be affected by checking. However, if the decision to stop depends on how precisely the difference is known, we no longer change the outcome by stopping while the test is still running. (Kovanen, 2017).

3.4.2 An alternative for traditional A/B testing through machine learning algorithms: multi-armed bandits

Automating manual work of A/B testing step by step as explained in the previous subchapter might offer great solutions for companies who still run their A/B tests in a traditional way. However, if we make space for alternative method for reaching the same goals, machine learning algorithms offer a solution. The multi-armed bandits (MAB) – problem is explained in previous research as follows. "In its simplest form, there are N arms, each providing stochastic rewards that are independent and identically distributed over time, with unknown means. A policy is desired to pick one arm at each time sequentially to maximize the reward. MAB problems capture a fundamental trade-off between exploration and exploitation: On the one hand, various arms should be explored in order to learn their parameters, and on the other hand, the prior observations should be exploited to gain the best possible immediate rewards" (Gai et al. 2012). As MAB-problems have been used widely in internet advertising optimization and network optimization (Agarwal et al. 2007), they can be seen as an alternative method for A/B testing, especially in SNSs. The possibilities that MAB enables for SNA optimization has not gained much focus in the research field even though the role of MAB as an alternative method for A/B testing has been discussed widely in the 2010s. Not having to go back continuously to check the effects of the tests and picking the options, and then iterating the tests again would save la lot of time. Letting the machine learning algorithm to constantly reveal the best alternative would release the companies from a lot of manual work regarding traditional A/B testing.

3.5 Practical insights from the case company

In the case company I conducted my question patter to, A/B testing is used in several activities: their digital display advertising channels, improving the usability of company's website, and optimization of sales. The display advertising channels where A/B testing is used include Facebook, real-time bidding, e-mail advertising and the company website. From these activities the channel that is most relevant and in the scope of my thesis is especially Facebook.

What it comes to the software that is used for analysing the result of the conducted tests, the company uses a specific program that is developed for A/B testing exactly that runs the tests, analyses the results and their statistical significance. This program, however, is used for the company website A/B testing. What it comes to the SNA context, the company collects and analyses data with Google Analytics and Excel. The company argues they use A/B testing for all kinds of objectives. For instance, in advertising, different combinations and pairings of text and picture alternatives are tested and also different landing pages. However, it is emphasized that in the company website everything from the site content (i.e. pictures and texts) to the functionality is tested through A/B. The case company has not used automation of A/B testing so far. However, they believe the automation will be the future direction as the need for user-specific optimization of the content increases.

I also found out more practical and manual aspects of the case company's A/B testing process including factors such as the sample size, amount of different variants, running time and tracking of the results. The tests are run with the maximum sample size – the entire reach of available users. The amount of variants, in turn, depends of the time when the test is conducted. The bottleneck is the amount of gained visits. If there are many variations, the views per variant –ratio can end up as a small number and thus the results may end up not statistically significant. In advertising A/B testing there are usually 2-5 variants and in the webpage usually two or sometimes more, according to the company's actions. The tests are run as long as the results are statistically significant, usually for a week. Sometimes, however, if the tested object is one that doesn't gain too many views, the test might be run for many weeks. In advertising the relevant result might be gained from many hours to a day. The results of the tests are tracked continuously so that the better version can be published as soon as the statistical significance is reached as the the final version.

4 Discussion

In this chapter I will summarize and conclude my key findings, describe the theoretical and practical implications of my work and analyse its limitations for future research. In my thesis I aimed to find answers to the limitations of traditional and current ways of conducting A/B testing in social network advertising and this way to find insights of the future development of it. As Magara and Tamburelli (2014) argue, the method of A/B testing as currently performed has been seen as a time consuming and error prone manual activity. Also, it has been characterized as a costly testing method. According to Thomke (2003) experimentation in general has been perceived demanding both time- and employee-wise. Of course, how A/B testing is considered and seen in the company depends a lot of its size, field, and marketing/ digital budget but according to my research there are ways to improve the A/B testing practice in SNA in most companies that are not widely adopted yet.

4.1 Key findings

The underlying problems regarding SNA A/B testing address employees' practical daily work and the definition and role of data in the organization (Magara & Tamburelli, 2014). Measurement in general as a problem solver should not be taken for granted in the process (Fink & Levenson, 2017). The useful or correct insights are not guaranteed even if measurement and experimentation are included in the processes and decision-making (Koomen & Siroker 2013). The employees working with the data collected through the A/B tests need to understand that this data includes not only accurate but also potential, irrelevant, and even flawed information. The rapidly increasing amount of data doesn't make this process easy for the employees and thus the data sources and their quality need to be challenged, as Fink and Levenson (2017) suggest. Careful definition of which parts of the A/B test results to include in the company decision-making and which to leave out is an important but sometimes underestimated or even ignored part of the process. For instance, the case company denotes they use A/B testing for all kinds of objectives. In this kind of cases it becomes increasingly critical to analyse the data carefully before making final conclusions from it. Furthermore, through the phenomenon of trusting the highest paid person opinion (HiPPO), the power of data can be underestimated in the organization (Henne et al. 2008). The aim of A/B testing in SNA is to gain insights from the customer preferences for the company's advertising optimization processes and hence improve profitability and return on investment (ROI). This occurs through the collection of information on the consumer needs and attitudes, data-driven decision making, and the careful consideration of such information, not blind-eyed focus on either data or anyone's personal opinions or views. Moreover, A/B testing includes practical work that easily results in flawed information gathering, increased time used for the tests and ineffectiveness in the usage of resources (Magara & Tamburelli, 2014). The automation of such manual tasks can be seen as a solution for these problems and results regarding technology used for the process automation is described later in this subchapter.

The other side of the SNA A/B testing besides the organization itself are the site users to whom the ads are targeted. User attitudes towards online advertising and their ad-clicking behaviour constitute another critical base for successful A/B testing. If user attitudes shift towards not preferring SNA at all, the entire process of SNA A/B testing will lose its purpose. On the other hand, SNA A/B testing can be seen as a method to influence the user preferences on whether viewing the ads or not. This relates strongly to the phase before the actual tests are run. The case company that contributed my work emphasizes the role of careful background work. This includes previous user experiences, indeed. Besides, previous tests and brainstorming different design variation prototypes add more to the background work. According to previous research by Baek and Morimoto (2012), and Lewin et al. (2011), the attitudes towards SNA are not affecting negatively the future of social network advertising. However, customers vary from individual to individual which leads to company having to interpret carefully the reasons for the lower click-through rates and reasons for the reasons for the one option losing the test. The relevance of these individual, heterogeneous aspects are crucial part of the future of SNA A/B testing as described next. Also, the case company insights support my findings from previous literature as the need for optimized content that varies from individual to individual is increasing.

Through the increasing focus on individual click-through rates, the possibilities of personalized visitor experiences and customized advertising are becoming more and more relevant. With the combination of both tracking the individual browsing paths and conducting careful A/B testing, each customer can be supplied with personalized ads, based on his or her individual preferences and the factors affecting their own ad-clicking only (Besbes et al. 2014). Besides, collecting information on for instance advertisement category, the timing, or the click history can lead to finding connections between the ad and the alternatives for the following recommendation, the ad-clicking behaviour, and the habits associated with the ads and the users clicking the ad. Thus, the companies can create paths to guide users from advertisement to another. It is no longer relevant to discuss one and specific website experience, as the same domain can show different users different personalized alternatives (Christian, 2012). However, the case company aspects indicate practical problems arising from the collection of this sort of detailed data required and utilization of it. They see automation as a solution for creation of such personalized visitor experience and customized advertising. Furthermore, the previous research doesn't indicate whether this kind of information on individuals could lead to privacy issues in the future or not. The questions on how personal consumers see their own individual ad-clicking paths, and how they feel about seeing different content than other users remain open.

As the A/B testing process requires a lot of both manual work in the forms of creating code and designing prototypes, and sophisticated skills such as statistical knowledge, automating the process and using machine learning algorithms should certainly be the direction of future development in the field (Magara & Tamburelli, 2014; Deng et al, 2016). Previous research has developed a few models for this purpose including the formulation of A/B testing as an optimization problem and hence implementing automated search algorithms (Magara & Tamburelli, 2016), automating A/B testing through Bayesian Hypothesis (Kovanen, 2017) and using multi-armed bandits in network optimization problems (Agarwal et al. 2007).

Considering the optional stopping problem (Kovanen, 2017), employees should not make real time data-analysis while the A/B test is running. The case company insights used in my data collection, however, indicate the company does check the results continuously. They tell they publish the "winner-option" immediately after the statistical significance is reached. If the optional stopping problem is not taken into consideration here, the results might end up flawed. This is an example of a situation where automation through Bayesian Hypothesis (Kovanen, 2017) would make relevant contribution to A/B testing. Besides, automation in general would increase the value of A/B test process in forms of being able to use more variants than they currently do. The company uses large sample size, in fact maximum, and a simple experimental design that are seen very important factors in the literature. Deng et al. (2014) suggest using, indeed, largest possible sample size to reach the optimal significance level and avoiding complex designs as those can hide errors.

The previous case is only one example of how all the companies using SNA A/B testing are not implementing automation or machine learning algorithms in their such processes even though it would release huge amounts of time, decrease the need for complex skills and reduce errors. This might be due to the lack of knowledge on the possibilities or ignorance of the importance of the topic. Moreover, the organizational capabilities might set the limits for modern technologies; having to educate employees to work with such technologies and practising change management. Besides, updating tools and software to support the methods might not be an easy task in every organization. This is what Donnelly and Durney (2012) describe as the inability of the regular management actions to work for rapid and complex technological change. They argue this sort of inability being due to the uncertainties that most leaders in such changing environments are not familiar with. However, according to the case company insights, the automated optimization will be in a bigger role. They already use some sort of advanced data tools for precisely A/B testing in their website optimization which could indicate that they have the interest and capacity to implement more sophisticated software also for their SNA. Besides, the case company emphasized the role of careful background work for the tests. If the process itself would be automated, more time for this kind of creative work would be available.

4.2 Implications to research

In this sub-chapter I assess my findings regarding their implications to previous research in more detail. My results mostly corroborate existing literature in the field. However, my results indicate that even though the A/B testing in social network advertising is widely known and well studied topic, it has not reached its full potential yet. According to the previous research, the knowledge and methods of developing the SNA A/B testing further do already exist but they are not as widely adopted as one might think. Especially the factors addressing the manual side of A/B testing that could more or less easily be overcome through automation, education and shift in the working habits, has not reached its full focus yet.

Besides, my work addresses the topic very broadly, taking large amount of perspectives into consideration collectively. Considering SNS A/B testing with the insights of the organizational capabilities on data and experimentation, user attitudes on SNA and their ad-clicking behaviour, personalized visitor experience, and automation of manual work and utilization of machine learning algorithms has not been summarized in the same paper. This might lead to new research directions that could open up new possibilities and transformation of current trends.

4.3 Implications to practice

Here I explain what the previously summarized results and key findings implicate in practice. I consider companies and other actors operating in the field and topic area of SNS A/B testing and give recommendations based on my findings.

What it comes to the organization's perceptions of the A/B test data, some general factors should be adopted within the employees. The workers in the advertising optimization should first understand the relevance of the data- and market-driven strategy in general. The careful implementation of experimentation culture should be implemented in the company to support this sort of mind set (Jenkins, 2014). Besides, the insights arising from the A/B test data should be challenged, analysed and criticized. Educating workers towards the data- and market-driven decision making and thinking can increase the value of the A/B testing process in the company which would result in successful advertising optimization and knowledge of the customer base and again in increased profitability and return on investment (ROI) (Leventhal, 2017).

The other critical side of A/B testing, the site users with varying attitudes and preferences, needs increased focus from the companies working on the tests. Due to the modern technology, it is possible to gather insights of individual users separately (Besbes et al. 2014). As the users vary from individual to individual in terms of their ad-clicking behaviours, it is recommended that this heterogeneity is included in the testing model (Duan et al. 2014). Moreover, the user attitudes and their ad-clicking behaviours should shift the direction of prototyping and other sort of background work, as the underlying attitude factors can already beforehand give insights on what kind of prototypes to create for the tests, even before they are started. Both viewpoints, organizational perceptions of data-driven business and user attitudes and their ad-clicking behaviour, increase the effectiveness of the A/B tests. These can be seen as factors that don't require rapid technological development but yet education and change in the organizational habits.

What it comes to the business benefits of creating individual ad-clicking paths and personalized visitor experience in terms of SNA, the companies should concentrate on collecting and analysing data on not only the ad-clicking behaviour, but also the contextual connections between ads and the visitors, patterns associated with the ads and the users clicking the ad, timely actions, etc. The creation of personalized visitor experience, in addition, might require utilization of A&B testing and Exceptional Model Mining (EMM) (Adegeest et al. 2017), as the traditional A/B testing most probably is not efficient enough for collecting data with the required volume.

The need for automation of manual work arises from the difficulties faced in the A/B test process. These processes include for instance creating, choosing and evaluating variants. The automation process requires not only resources and capacity for the big investment. Updating company's software and tools to support the new methods might cost a lot of money but also updating the skills of the employees working with the tests requires attention. Donnelly and Durney (2012) suggest managers to focus on things such as clearly articulating the risks of the project, defining each individual's roles, conducting flexible management style, and finally, encouraging culture that supports innovation. This kind of change management for technological transformations, for instance, reduce change resistance in the organization. Besides practical education on how the work image changes from manually conducting the tests to guiding the automated process depends on the level of automation.

4.4 Limitations and future research

The limitations of this study lie in its methodology and data in general. First, literature review is a limited way of conducting research on future development of A/B testing as the technology might not yet exist and my research doesn't indicate how well the suggestions work in practice. Second, as the practical insights are collected from one specific case company, they cannot be generalized across countries, industries, user groups or time periods. Third, as I scoped my research to address only A/B testing in business processes, in fact only marketing and again only advertising context, the results cannot be straight forwarded to any other fields of science.

Propositions to future research include empirical research on how companies on a large scale actually perform A/B testing in SNA. Empirical data for example on what are the key metrics, how sensitively they react on changes on key metrics, which factors affect the creation and chose of advertising prototypes, how long they run the tests, and how quickly their developers realize performance would be interesting to collect and analyse. Suggestions for such questions have been already been made by Deng et al. (2014) but there's no indication if companies actually perform their A/B tests with these research guidelines in mind. These factors could then be compared to the existing theories, and more specific, concrete and credible future suggestions could be made. Besides, it would be interesting to research how large percentage of companies actually perform automated A/B testing and at which level the automation occurs (Magara & Tamburelli, 2014). Lastly, I found no previous research on the privacy issues arising from collection of individual user data through A/B testing and making conclusions based on those. I believe that as the processes of A/B testing in SNA will become more and more multidimensional, the roles of personal and privacy issues will increase. Moreover, the results from the case company indicate they use a sophisticated software program that is available in the market for collecting and analysing data of company website A/B testing. Otherwise they use Google Analytics and Excel. This in mind, it would be very interesting to research what kind of software programs are used in the SNA A/B testing process in different sort of companies and which A/B test activities they prioritize over SNA A/B test programs budget-wise.

5 Conclusions

My thesis focuses on the future of A/B testing in social network advertising. I have covered the subject from the perspectives of organizational capabilities and knowledge on utilizing data, user attitudes and their ad-clicking behaviour, personalized visitor experience, and automation through machine learning algorithms. According to this literature review, the focus should be shifted to careful consideration of both the A/B testing data and the user attitudes towards the social network advertising, the errors and bad utilization of resources arising from crafted work image of conducting A/B tests, and potential investments on more sophisticated software programs that help in automating the process or parts of it. These occur through understanding the full potential of SNA A/B testing, encouraging employees towards experimentation, and data- and market-driven decision-making, careful management of technological change and implementation. If these factors are overcome, the future of A/B testing is going to shift towards the more and more comprehensive automation of the processes that yet are very manual and practical. This way, more time-efficient, error-free, and multidimensional results can be gained through the SNS A/B testing. Moreover, organizational HR capacities can be released as the workers could reduce their time used for precise tasks and instead focus on innovation and creation of A/B test prototypes and other background work, and utilization of the entire process. Moreover, the creation of ad-clicking paths and personalized visitor experience will most likely, with the process automation and increasing technological investments, lead the way as the future direction of SNA A/B testing.

References

1. Adegeest, J & Duivestein, W. & Farzami, T. & Foks, G. Pechenizkiy, M. & Peer, E & Putman, T. & Weerts, J.P. 2017. *Have It Both Ways — from A/B Testing to A&B Testing with Exceptional Model Mining*. ECML PKDD 2017
2. Argawal, D. & Chakrabarti D. & Pandey, D. 2007. *Multi-armed bandit problems with dependent arms*. International Conference on Machine Learning, 24 pp. 721–728.
3. Backstrom, Lars & Kleinberg, Jon. 2011. *WWW 2011 – Session: Social Network Algorithms*. pp. 615-624.
4. Baek, Tae Hyun & Morimoto, Mariko. 2012. *Stay Away from me: Examining the determinants of consumer avoidance of personalized advertising Journal of Advertising*, 41 (1). 59-76.
5. Besbes, O. & Gur, Y. & Zeevi, A. 2014. *Optimization in Online Content Recommendation Services: Beyond Click-Through Rates*. Manufacturing & Service Operations Management, 18 (1), pp. 15-33.
6. Cho, C.H. 2003. *Factors influencing clicking of banner ads on the WWW*. Cyber Psychology & Behavior. 6 (1). pp. 201-215.
7. Christian, Brian. 2012. *The A/B Test: Inside the Technology That's Changing the Rules of Business*. The Wired Magazine. Accessed December 14, 2017 at http://www.wired.com/business/2012/04/ff_abtesting
8. Cubero, R. A. & Koukouvis, K. & Pelliccione, P. 2016. *A/B Testing in E-commerce Sales Processes*. Lecture Notes in Computer Science, 9823. pp. 134-148.
9. Deng, A. & Chen, S. & Lu, J. 2016. *Continuous Monitoring of A/B Tests Without Pain: Optional Stopping in Bayesian Testing*. IEEE International Conference on Data Science and Advanced Analytics.
10. Deng, A. & Kohavi, R. & Longbotham, R. & Xu, Y. 2014. *Seven rules of thumb for web site experimenters*. 1857-1866. ACM Lawler, Edward E. Doing research that is useful for theory and practice 1985. Lexington Books.
11. Denney, Andrew S. & Tewksbury, Richard. 2003. *How to Write a Literature Review*. Journal of Criminal Justice Education, 24 (2).
12. Donnelly, Richard G. & Durney, Christopher P. 2012. *Managing the Effects of Rapid Technological Change on Complex Information Technology Projects*. Journal of the Knowledge Economy, 6 (4). pp. 641-664.

13. Duan, J.A. & Whinston, A. & Xu, L. 2014. *Path to purchase: A mutually exciting point process model for online advertising and conversion*. Management Science, 60 (6), pp.1392-1412.
14. EY. 2017. *Becoming an analytics-driven organization to create value*. Accessed December 12, 17 at http://www.ey.com/Publication/vwLUAssets/EY-global-becoming-an-analytics-driven-organization/$FILE/ey-global-becoming-an-analytics-driven-organization.pdf
15. Fink, Alexis & Levenson, Alec. 2017. *Journal of Organizational Effectiveness: People and Performance Human capital analytics: too much data and analysis, not enough models and business insights*, 4 (2). 145-156.
16. Gai, Y. & Jain, R. & Krishnamachari, B. 2012. *Combinatorial Network Optimization with Unknown Variables: Multi-Armed Bandits with Linear Rewards and Individual Observations*. IEEE/ACM Transactions on Networking, 20 (5). 1466-1478.
17. Geng, Z. & Glass, A. & Jiang, B. & Shang, H. & Shi, X. 2016. *A Framework for Network A/B Testing*. eprint arXiv:1610.07670. Accessed November 8, 2017 at https://arxiv.org/pdf/1610.07670.pdf.
18. Golsby-Smith, Tony & Martin, Roger L. 2017. *Management is much more than a science: the limits of data-driven decision making*. Harvard Business Review, 95 (5) pp. 128, Harvard Business School Publishing, United States.
19. Henne, R. & Kohavi, R. & Longbotham, R. & Sommerfield, D. 2009. *Controlled experiments on the web: survey and practical guide*. Data Mining and Knowledge Discovery, 18 (1).
20. Jenkins, Wyatt. 2014. *A/B Testing and the Benefits of an Experimentation Culture 2014*. Harvard Business Publishing, Harvard Business Review, February 5th 2014.
21. Kovanen, Lauri. 2017. *Statistical Significance for Humans – Automated Statistical Significance Calculator for A/B Testing*. Accessed December 15, 17 at https://www.smartly.io/blog/statistical-significance-for-humans-automated-statistical-significance-calculator-for-a/b-testing
22. Leventhal, Richard C. 2005. *The importance of marketing Strategic Direction*. Emerald Group Publishing Limited, 21 (6). pp. 3-4.
23. Lewin, J. E. & Strutton, D. & Taylor, D. G. 2011. *Friends, fans, and followers: do ads work on social networks?* Journal of Advertising Research, 51 (1) pp. 258-275. Journal of Advertising Research.

24. Magara, Giordano & Tamburelli, Alessandro. 2014. *Towards Automated A/B Testing*. International Symposium on Search-Based Software Engineering, 184. Springer Verlag.
25. Malthouse, Edward C. & Li, Hairong. 2017. *Opportunities for and Pitfalls of Using Big Data in Advertising Research*. Journal of Advertising, 46 (2). pp. 227-235. Taylor & Francis.
26. Mir, Imran. 2015. *Effects of Beliefs and Concerns on User Attitudes towards Online Social Network Advertising and Their Ad Clicking Behavior*. Journal of Internet Banking and Commerce, 20 (2).
27. Mohrman, Susan A. & Lawler, Ed. 2011. *Useful research: Advancing theory and practice*. Berrett-Koehler Publishers.
28. Negnevitsky, M. 2004. *Artificial intelligence: A guide to intelligent systems*. 2nd ed. Addison-Wesley.
29. Siroker, Dan & Koomen, Pete. 2013. *A/B Testing: The Most Powerful Way to Turn Clicks into Customers*. John Wiley & Sons Incorporated. ProQuest Ebook Central, 1.
30. Thomke, Stefan H. 2003. *Experimentation matters: unlocking the potential of new technologies for innovation*. Harvard Business Press.
31. Young, Scott. 2014. *Improving Library User Experience with A/B Testing: Principles and Process*. Journal of Library User Experience, 1 (1).