The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2040-7122.htm

Impact of direct traffic effect on online sales

Lukáš Kakalejčík
Department of Applied Mathematics and Business Informatics, Technicka univerzita v Kosiciach, Kosice, Slovakia
Jozef Bucko
Department of Applied Mathematics and Business Informatics, Faculty of Economics, Technical University of Košice, Kosice, Slovakia, and
Jakub Danko
Department of Statistics and Probability, University of Economics, Prague, Czech Republic

Received 27 January 2019
Revised 7 August 2019
Accepted 28 January 2020

Abstract

Purpose – This study aims to analyze the impact of newly created brand awareness on customer's buying behavior in online environment.

Design/methodology/approach – The authors analyzed more than 280,000 online customer journeys from four e-commerce stores based in Slovakia. Within the results of the interaction analysis of individual customer journeys, the authors determined three criteria based on the level of theoretical brand awareness. The purpose was to determine their occurrence in real-world data.

Findings – It was found that each of the specified criteria accounts for the significant share of the company's revenues. Based on these criteria and the level of their occurrence, the authors introduced the term direct traffic effect.

Research limitations/implications – Because of the available Web analytics tools, the data might be imprecise because of data collection issues. There is also ambiguity in the interpretation of the customer journey.

Practical implications – The company can build awareness among prospective customers by offering them a positive customer experience during the first interactions online. Data proved that customer will not only repeatedly visit the website from the direct traffic source but also his customer journey will end with the purchase of the company's products.

Originality/value – This paper fulfills the need for further research on the impact of multi-channel marketing on brand awareness and consumer behavior, respectively.

Keywords Online consumer behaviour, Marketing research, Customer analytics, Multi-channel measurement

Paper type Research paper

Introduction

In 2004, a new stage of the Web environment began and the classic solid content of the websites was replaced by a space for sharing and joint content creation (Dinucci, 1999). This development stage, known as Web 2.0, plays an important role in the functioning of businesses. By eliminating time and space boundaries, it gives the business the ability to reach out to its potential customers anywhere and anytime in real-time. The result of this opportunity is a partial or complete shift of business focus to the internet or mobile applications. Furthermore, internet users' movement is well-measurable, allowing

Journal of Research in Interactive Marketing
© Emerald Publishing Limited
2040-7122
DOI 10.1108/JRIM-01-2019-0012

companies to analyze not only engagement, transactions and customer sales but also allows them to analyze their potential customers – the source of their visits, their behavior, and also how far they are to become customers of the company (Clifton, 2015).

More than 90 per cent of users are not ready to buy during their first visit to the business website (Vanden Heuvel, 2014). On the contrary, from the first visit until the purchase, the users go through a process called the customer journey. The customer journey represents the sequence of steps that users gradually pass through the awareness stages, evaluating alternatives up to the actual purchase of the product (Roberge, 2015). Customer journey mapping is a model, which describes all interactions with the intent to improve these interactions, resulting in an increase in sales and customer satisfaction (Van Den Berg and Pietersma, 2015). With digital advertising booming in rising amount of platforms and formats, it is even more important to track consumer's digital footprints at a detailed level, enabling advertisers to get deeper insights into online consumer's behavior, as well as an image of what is the impact on conversion in case of the exposure of customer to individual advertising channels (Ghose and Todri, 2015). As customers do not live in a silo, their customer journey consists of interactions with both online and offline marketing channels. This paper focus on the online part of the customer journey as online channels are considered to be suitable channels for sales activation and with the current development of analytical tools, it is possible to attribute the value to particular online channels (Binet and Carter, 2018). The following marketing channels are among the most frequently measured and reflected by Web analytics tools:

- *Direct traffic*: represents a situation in which a user enters the URL of a website directly into the browser window or he visits a website through a saved bookmark. Visits from mobile apps or offline advertising sources (TVs, billboards, flyers, etc.) may also be considered as a direct visit in case of inappropriately selected or implemented tracking of traffic sources;
- *Organic traffic*: represents a situation where a user enters a key phrase in the search engine (Google, Bing, Yahoo and others), and clicks on search results to go to the website of the business;
- *Referral traffic*: represents a user's visit by clicking on a link placed on another website (it usually does not include social networks);
- *Social media*: represents a user's visit by clicking a link placed on social networks (Facebook, Twitter, LinkedIn and others);
- *E-mail*: represents the user's visit by clicking the link in the e-mail delivered to his mailbox;
- *Paid search*: represents the user's visit by clicking on paid search results (e.g. Google Ads platform); and
- *Display advertising*: represents the user's visit by clicking on a banner ad placed, for example, on the Google Display Network; and other, less frequently used marketing channels for online promotion.

The increase in potential customer touchpoints and the reduced control of the experience require firms to integrate multiple business functions to create and deliver a positive customer experience (Lemon and Verhoef, 2016). Usually, businesses do not use only one marketing channel to get the customer. These channels, in most cases, work in a cohesive way that contributes to the customer's acquisition. A merit value for acquiring a customer should be assigned to each such channel. Attribution models are used to model this problem. Attribution modeling (multichannel attribution) is a set of rules that give the individual marketing channel credit for obtaining a customer conversion (Shao and Li, 2011; Clifton, 2015). Danaher and Van Heerde (2018) define attribution modeling as the science of using advanced analytical methods to allocate sufficient credit for each contact/interaction of the customer with marketing channels used by the business. In the previous studies that we have carried out (Ferencová *et al.*, 2015), a problem has been defined with the evaluation of the utility of marketing channels during the sales cycle. In spite of the questionnaire survey conducted on customers, it is often difficult for the customer to determine which channels he interacted with prior to the purchase. This problem can be solved by attribution modeling, which evaluates every customer interaction with the business. However, in this article, attribution modeling is essential only to understand the complexity of the customer journey in terms of interacting with the company before purchasing the product.

There have been several studies that offered data-driven approaches to the attribution to overcome the weaknesses of standard heuristic models. Yadagiri *et al.* (2015) and Nissar and Yeung (2015) use Shapley value in their non-parametric approach to attribution as a game theory-based model. In his thesis, Rentola (2014) used two models: binary logistic regression to classify customers to converters and non-converters (purchasers/non-purchasers), as well as a logistic regression model with bootstrap aggregation. On the other hand, Shao and Li (2011) used bagged logistic regression and a probabilistic model in their study. In their study, Li and Kannan (2014) used a hierarchical Bayesian model. Geyik *et al.* (2014) developed their attribution algorithm MTA to solve two problems: spending capability calculation for a sub-campaign and return-on-investment calculation for a sub-campaign [more in (Geyik *et al.*, 2014)]. On the contrary, Wooff and Anderson (2015) offer an attribution mechanism based on the appropriate time-weighting of clicks using the sequential analysis. Hidden Markov model was used in the studies conducted by Abhishek *et al.* (2012) and Wang *et al.* (2015). Markov chain model was proposed in several studies as well (Anderl *et al.*, 2014; Anderl *et al.*, 2015, 2016). For the purpose of our study, we adopted the Markov model with the GDL estimator used in the study by Kakalejčík *et al.* (2018) to determine the importance of online marketing channels.

Previous studies on attribution modeling (Anderl *et al.*, 2016; Rentola, 2014; Li and Kannan, 2014) have shown that a specific source – *Direct Traffic* – has been an important marketing channel with the merit of generating purchases and sales. *Direct Traffic* can be labeled as a brand awareness aspect. Brand awareness could be defined as the extent to which consumers are familiar with the distinctive qualities or image of a particular brand of goods or services. Awareness is distinguished in terms of two dimensions as follows: intensity and extent. The intensity of brand awareness indicates how effortlessly consumers recall a particular brand. The extent of brand awareness refers to the possibility of acquiring and consuming brand services and products especially when the brand emerges in consumers' minds (Barreda *et al.*, 2015). The extent of brand awareness can be understood as physical availability defined by Romaniuk and Sharp (2015), which in the context of online media (such as websites or social networks) refers to the possibility to purchase the product online. Brand awareness, in accordance with established marketing theory and practice standards, can take three forms:

(1) *Top-of-mind*: represents the gold standard of brand awareness. In this case, the brand is the first to be remembered by the customer without any help/support;

(2) *Spontaneous awareness*: represents recognition of the brand by the customer, with no help given; and

(3) *Supported awareness*: different help is given to the customer, and it is monitored, which brands come to his mind in the given context (Kahn, 2013).

Brand awareness plays a key role in consumers' buying decision-making process (Binet and Carter, 2018). It possesses aspects such as individual recognition, the dominance of knowledge and brand recall (Kim *et al.*, 2008). When typing the name of the website to the Web browser window, the customer usually have to type the brand name, and therefore, remember or recall it. Consequently, *Direct Traffic*, in most cases, might be connected to spontaneous awareness or even "top-of-mind" brand awareness – depending on whether the website came to mind of the customer as the first or another alternative to solve his current problem. Ash (2012) claim that *Direct Traffic* referral means that the person is specifically aware of and looking for the company. It is usually achieved as a result of repeated exposure to the brand's diverse settings. Bones *et al.* (2019) discuss that the analysis of *Direct Traffic* over time can help brands understand changes in brand awareness, especially rises in *Direct Traffic* can provide an indication of increased brand awareness. In the context of brand awareness building in the online environment, Barreda *et al.* (2015) discuss that the quality of the system (navigation simplicity, good user experience and security), the quality of information (information that help the user to make a better decision), rewards (users obtain financial, psychological or membership privileges) and virtual interaction (the range in which users can participate in changing the content of the website in real-time) are among the bearers of building the brand awareness. This means that the user is transitioning from the search of how to resolve the problem to a direct business visit, which can solve his problem under the conditions of meeting these prerequisites through the business website.

Anderl *et al.* (2015) divide the sources of website visits into two categories as follows:

(1) Channels initiated by the customer, which are further divided into branded (including *Direct Traffic*) and generic;

(2) Business-initiated channels: it represents the promotional activities of the business (*e-mail, affiliate, banner ads*, etc.).

Anderl *et al.* (2015) in their taxonomy model describe that after several visits to the website through the business-initiated channels, the customer moves to the stage when the visits are initiated by himself, which is a shift in customer decision-making about the product during the purchasing process. The shift that has just been mentioned, is the subject matter of this part of the paper. Li and Kannan (2014), in this context, list the concepts of spillover effect and carryover effect. A *carryover effect* occurs when a user visits a website through a single source of traffic, and subsequently, visits (and eventually) buys during a visit from the same source. On the contrary, a *spillover effect* occurs when a user visits a website through a single source of traffic, and subsequently, visits (and eventually) buys during a visit from another source. Within our analysis, we will assume that both these events will occur in the context of *Direct Traffic* as a brand awareness indicator. Linking brand awareness to the source *Direct Traffic* is the research content of the presented study.

Sample and methods

The presented study aims to analyze the impact of online brand awareness on customer's purchasing behavior, based on the current state of knowledge. By decomposition of this objective, we determined the following partial objectives:

- The analysis of the current state of discussed issue in areas of multichannel attribution and online brand awareness;

- Determine the importance of the source *Direct Traffic* by using Markov chains for multichannel attribution and analyzing transition probabilities, as well as removal effects in case of four Slovak e-commerce stores and compare it with the results from the previous studies (Anderl *et al.*, 2016; Rentola, 2014; Li and Kannan, 2014); and
- Observe the incidence of three defined criteria connected to online brand awareness and determine their impact on the business performance of the analyzed e-commerce stores.

Primarily, we focused on "website stickiness" after first customer visits, based on the taxonomy model defined by Anderl *et al.* (2015). For the purpose of our study, the data from four e-shops (companies) were collected, one of which is focused on the sale of electronic components, two of which are focused on the sale of sportswear, the latter focusing on the sale of nutritional supplements [Table I displays characteristics of businesses based on Finstat (2017)]. The names of the companies remain anonymous in agreement with their representatives. As was mentioned in the introduction of this paper, customers do not live in a silo and interact with both offline and online communication activities of the companies. To prevent offline communications to have an effect on our results, we selected companies that did not execute any offline communication three months prior to and during the period of data collection.

The data on customer journeys of e-shop customers of the analyzed companies were obtained from the Google Analytics (2018) platform that companies use to measure the performance of their websites (e-shops). The description of input data in terms of the number of customer journeys and the volume of generated revenue is shown in Table II.

To perform the analysis, the following customer journeys were excluded from the available customer journeys data sets in the first step:

- Customer journeys beginning with *Direct Traffic* that indicate previous brand awareness or brand experience (re-purchase, offline promotion of the business – are in contradiction with study goals). There are several limitations to this step: when inactive cross-device tracking of users, the customer can initiate the exploration phase on the mobile and later purchase the product on the desktop device by remembering the URL of the page. In this case, in Web analytics software, the customer journey starts with *Direct Traffic*; users and their customer journeys are tracked through cookies. If a customer commenced a product survey, deleted his cookies and then purchased the product by visiting the website after entering the URL address in the browser, the shopping journey recorded by the analytics software starts with *Direct Traffic*; in customer journeys, it was not possible to mark search queries from organic search related to the brand as *Direct Traffic*. So, if a user searched for a store name through a search engine, visited the website and later purchased, his customer journey starts with the organic search source instead of *Direct Traffic*; and
- Customer journeys that do not contain the *Direct Traffic* source.

Subsequently, three monitored criteria were defined, on the basis of which we conducted the analysis:

- *Criterion 1*: the last source (step) in the customer journey is *Direct Traffic*. In this case, we do not monitor the spillover or carryover effect because both cases may occur (customer journey may end *Direct Traffic* > *Direct Traffic* and *Social*

Table I.
Characteristics of the analyzed companies

	Company 1	Company 2	Company 3	Company 4
Subject of activity	Distribution of industrial electronic components for industrial production	Retail sale of sporting goods of a wide range	Retail sale of sporting goods with a focus on running and triathlon	Retail sale of food and nutritional supplements
Revenues in 2016 in this €	15,561	16,018	308	4,993
Number of employees	50-99	200-249	3-4	20-24
Monitored period of customer journeys	April 1, 2016- August 31, 2016	July 1, 2016-June 30, 2017	July 1, 2016-June 30, 2017	December 4, 2016-December 4, 2017

networks > *Direct Traffic*). We are only interested in the impact of *Direct Traffic* (and therefore, brand awareness) on the buying behavior of the users;

- *Criterion 2*: in the customer journey, the *Direct Traffic* source is at least twice in a row at any point in the customer journey. In this case, we monitor the spillover effect from another channel into the channel we are monitoring, as well as the transfer effect in the *Direct Traffic* source that indicates brand awareness. In this case, we are not interested in what channel the customer journey ends with. Anderl *et al.* (2015) discuss that there is a positive interaction effect when using customer-initiated channels and the following business-initiated channel visits, which is greatly influenced by the company's ability to use remarketing strategies. Remarketing (also referred to as behavioral targeting) is a way to promote a product to the people who have previously visited a business website (Marshall and Todd, 2017). In the case of a user's reaction to a remarketing campaign, the customer journey can end with, for example, the source banner ad or paid search; and
- *Criterion 3*: in the customer journey, the *Direct Traffic* source is at least three times in a row at any point in the customer journey. This number indicates even stronger brand awareness than Criterion 2. At the same time, it represents either the formation of positive brand preferences or the user's decision-making among the available alternatives to the product offered by other businesses as well. In this case, we again monitor the spillover effect from another channel into the channel being monitored by us, as well as the transfer effect in the *Direct Traffic* source that indicates brand awareness. As in the previous case, we are not interested in what channel/source of traffic the customer journey ends with.

Before the analysis of the criteria was conducted, we used Markov chains to determine the importance and value of *Direct Traffic*. Formally, a sequence of random variables $\{X_t\}_{t=1}^{\infty}$, $X_t \in S := \{s_1, \ldots, s_m\}$, is a Markov chain of order r if, for all $(a_1, \ldots, a_{t+1}) \in S^{t+1}$, $P(X_{t+1} = a_{t+1} \mid X_1 = a_1, \ldots, X_t = a_t) = P(X_{t+1} = a_{t+1} \mid X_{t-r+1} = a_{t-r+1}, \ldots, X_t = a_t)$ and r is the smallest integer to satisfy it. Essentially, this represents that the probabilities related to X_{t+1} depend only on the last r events, for all t.

In this context, S is referred by the state space, a particular sequence $(a_1, a_2, \ldots) \in S^{\infty}$ is called by a trajectory, the size of S is the length of state space or number of states, represented by m, and the probabilities of $X_{t+1} = a_{t+1}$ considering that $(X_{t-r+1}, \ldots, X_t) = (a_{t-r+1}, \ldots, a_t)$ are called the transition probabilities represented by the notation $p(a_{t+1} \mid a_{t-r+1}, \ldots, a_t := P(X_{t+1} = a_{t+1} \mid X_{t-r+1} = a_{t-r+1}, \ldots, X_t = a_t)$. A particular state b is absorbing if the probabilities to leave the state are "0", i.e. $p(c \mid a_{t-r+1}, \ldots, b) = 0, \forall\, c \neq b$, and consequently, $p(b \mid a_{t-r+1}, \ldots, b) = 1$.

A Markov chain can be represented by an initial probability distribution for the first r steps and the m^{r+1} transition probabilities. When $r = 1$, it is possible to have a graphic

Table II. Characteristics of input data on customer journeys

	Company 1	Company 2	Company 3	Company 4
Number of conversions	6,304	21,119	2,118	255,034
Total amount of purchases	€1,579,778.00	€976,514.82	€219,719.49	€4,889,682.05
Average order value	€250.60	€46.24	€103.74	€19.17
Customer journey duration (mean)	23.92	15.72	15.73	20.20
Customer journey duration (median)	14	9	6	10

representation for the Markov chain. For more details about Markov chains, we recommend (Karlin and Taylor, 1975).

Anderl *et al.* (2014) propose the use of Markov chains on channel attributions, considering the state space S as the states "start" and "conversion" combined with the set of marketing channels. In this case, the process $\{X_t\}$ represents the possible customer journeys through these channels. They suggest using a removal effect for attribution modeling. The removal effect is defined as the probability to achieve the conversion from the "start" state if some of the states (s_i) are removed from the model. As the removal effect reflects the change in conversion rate if the given state s_i is removed, the value (or importance) of the given marketing channel can be determined. If N conversions are generated without the particular channel (compared to the number of conversions in the full model), the removed channel determines the change in the total number of conversions (Bryl, 2016).

In addition to the above methods and procedures, elements of descriptive statistics and characteristics of variables (average, median and quartiles) were used for data analysis. Data were analyzed using the statistical platform The R Project for Statistical Computing (2016).

Results

To vindicate the importance of this study, it was necessary to examine the importance of *Direct Traffic* and its impact on sales. In the first part of the study, we analyze all of the buyer journeys with the use of Markov chains in accordance with the theory used in the previous section of the paper. We are particularly looking at the transition probability from any other channel to *Direct Traffic* and the transition probability from *Direct Traffic* to purchase. In addition, we are interested in measuring the removal effect, which is the direct indicator of the importance of the marketing channels. Afterward, we proceeded with the analysis based on the three criteria we determined in the methodology section of this paper.

The initial step of the analysis was the generation of transition diagrams and also the generation of transition matrices for each analyzed company. Figure 1 shows the customer journeys transition diagrams for all the analyzed businesses. The individual points of the graphs represent specific states – marketing channels. It can be noticed that the transition diagram starts with the state (*start*) that represents the start of the customer journey, and ends with the state (*conversion*) that represents the conversion or transaction. Individual states are linked by nodes, each node containing information about the transition probability from a particular state to another particular state. The nodes between the two marketing channels m and n show two probabilities – the probability of transition from state m to state n and the probability of transition from state n to state m. The nodes that connect states (*start*) and (*conversion*) contain only one probability because no customer journey is heading to the state (*start*). Likewise, no customer journey is heading from the state (*conversion*) toward the marketing channels used. This is a logical state because, after the performed transaction, it does not make sense to monitor what marketing channels the customer uses at the point of further interaction with the business. In the state (*conversion*), you can see a loop with a pictured probability of 1. This loop originated from computational reasons during the implementation of all the customer journeys. As all the customer journeys have to go from the state (*start*) to the state (*conversion*), the loop serves to complete each further interaction until the moment when all the customer journeys from the data file get through the state (*conversion*).

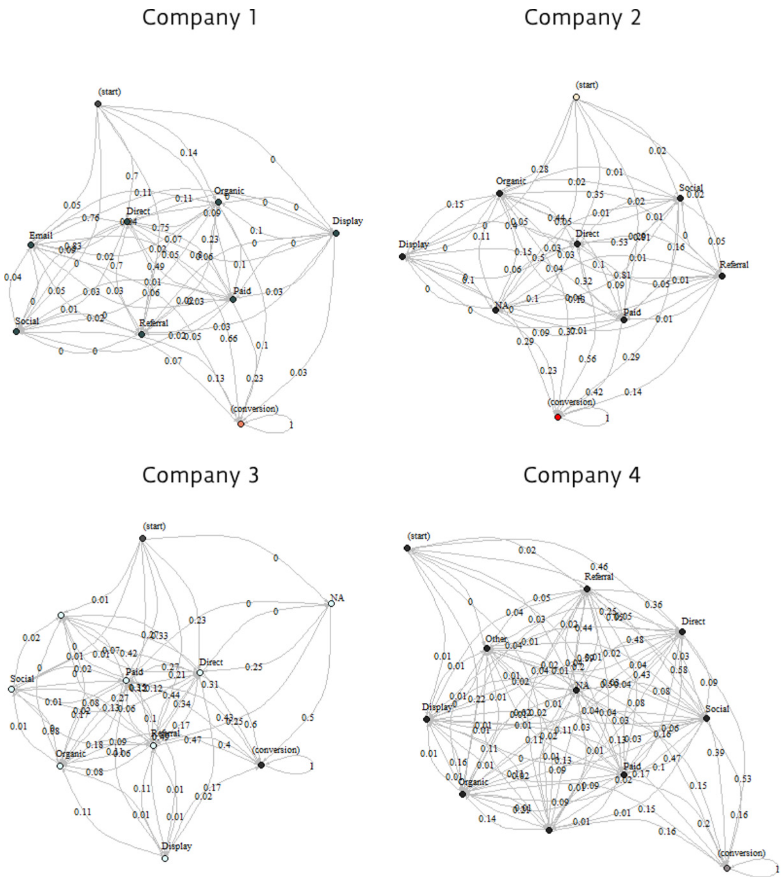

Figure 1.
Transition diagrams of the customer journeys

When analyzing the transition probability, two events can be observed in all four companies:

(1) If we do not consider completing a purchase (conversion), with almost every marketing channel, the most likely next step is to visit the website from the *Direct Traffic* source. This means that customers will either remember the page (and will write a URL directly to the browser on the next visit) or they will save the page as a bookmark to access it later; and

(2) When visiting a website whose source is labeled as *Direct Traffic*, there is the highest probability that customers will purchase the product.

When analyzing the selection of the appropriate order for the Markov chain, changes in the removal effect were also in the center of our attention. It is established that the higher the removal effect of a given marketing channel, the more important a marketing channel for the business because excluding it from the marketing portfolio would greatly reduce the number of transactions (conversions) achieved. Table III shows the removal effects using the first-order of the Markov model in terms of conversions (C), as well as in terms of revenue generated (R).

From Table III, it is obvious that *Direct Traffic* is reaching the highest removal effects, both in terms of conversions and revenue. This knowledge directly supports the findings from the previous part of the analysis, which concluded that visits from the *Direct Traffic* channel have the greatest chance of ending up with a purchase. These findings are corresponding with the results obtained in previous studies (Anderl *et al.*, 2016; Rentola, 2014; Li and Kannan, 2014). These results lead us to analyze the impact of *Direct Traffic* on sales more in detail – by examining results in accordance with the criteria set in methodology.

Table IV illustrates the occurrence of the given criteria in the customer journeys of the analyzed companies. It can be noticed that the tightening of the criteria also decreases the share of the number of customer journeys in their total number. When comparing Criterions 1 and 3, the number of customer journeys decreases in all analyzed companies by more than a half. The results of the analysis carried out are of the highest importance for Company 4 because, in both absolute and relative numbers, customer journeys according to established criteria occur the most in case of this company in comparison with other analyzed companies.

Table V shows the relative frequency of the number of interactions (website visits) that precede a visit from the *Direct Traffic* source and end with a purchase (Criterion 1). Looking at the table, it is possible to see that the values of the four analyzed companies are very similar. Approximately 30 per cent of the Criterion 1 customer journeys require only one prior visit from another source to follow a visit from the *Direct Traffic* source that ends with the purchase of the product. This phenomenon can represent a high "likeness" of a product or more precisely brand that results in spontaneous brand awareness and subsequent purchase with a low number of visits to the website (the entire customer journey can also

Table III.
The number and proportion of customer journeys according to established criteria

	Company 1	Company 2	Company 3	Company 4
Total number of conversions	6,304	21,119	2,118	255,034
Criterion 1	1,310	5,347	465	80,271
Criterion 1 (%)	20.78	25.32	21.95	31.47
Criterion 2	880	3,402	259	51,497
Criterion 2 (%)	13.96	16.11	12.23	20.19
Criterion 3	632	2,284	167	34,656
Criterion 3 (%)	10.03	10.81	7.88	13.59

Table IV.
Removal effects (Markov model of the first-order)

	Company 1		Company 2		Company 3		Company 4	
	C	*R*	*C*	*R*	*C*	*R*	*C*	*R*
Direct traffic	0.95	0.92	0.70	0.73	0.58	0.61	0.86	0.89
Organic search	0.30	0.32	0.39	0.38	0.40	0.37	0.41	0.40
Reference resources	0.10	0.12	0.17	0.17	0.36	0.40	0.13	0.14
Social networks	0.02	0.02	0.06	0.06	0.03	0.03	0.21	0.21
E-mail	0.16	0.17	–	–	0.02	0.02	0.06	0.05
Paid search	0.21	0.22	0.41	0.40	0.38	0.36	0.41	0.39
Banner advertising	0.01<	0.01	0.01	0.01	0.02	0.02	0.03	0.03
Other	–	–	–	–	–	–	0.03	0.03
N/A	–	–	0.18	0.18	0.01<	0.01<	0.11	0.11

last only a single day). Relative frequency of the number of interactions from 1 to 4 interactions prior to the *Direct Traffic* source visit decreases by 30-40 per cent with each additional interaction. Concerning the users who end up their customer journey by purchasing while visiting a website from the *Direct Traffic* source, it is possible to see high decisiveness. This is also evidenced by the cumulative proportions of the length of customer journeys for each analyzed company, which are shown in Figure 2. It says that approximately 90 per cent of the customer journeys are composed of 10 or fewer interactions. However, as the last interaction from the *Direct Traffic* source is not included in this number, this statement needs to be modified as follows: approximately 90 per cent of customer journeys ending in a visit from the *Direct Traffic* source consists of 11 or fewer interactions. The conclusion is that brand awareness has a positive impact on shortening the customer journey before the purchase made by customers online.

The goal of Criterion 2 was to analyze the number of interactions with a website that triggers spontaneous brand awareness so that users (customers) visit the website twice in a row through the *Direct Traffic* source, regardless of what source they make the purchase itself when visiting. Thus, the customer journeys can end with the *Direct Traffic* source but they can also end with any other source from the spectrum used. Table VI states that in customer journeys where this criterion was met, 66-72 per cent of customers had only one visit to build relatively strong brand awareness, they were in more frequent interaction with the website of the companies and completed their customer journey with a purchase. The conclusion is that in the online environment, a part of the customers' needs just one interaction with the website to remember the name of the brand or the URL of the website, to buy on this website later. However, the number of these purchases is low, as the monitored customer journey must consist of a specific number of interactions so that the set criterion can be reached.

As mentioned, very strong brand awareness is expressed by set Criterion 3. Within it, customer journeys have been analyzed, showing a visit from the *Direct Traffic* source three times in a row, which means a very strong engagement and absolute brand awareness or website as such. Looking at Table VII it can be noticed that the values of 1 and 2 of the previous interactions are almost identical to those of Criterion 2 (however, they are slightly lower with the exception of Company 1 concerning the two previous interactions). However, from the number of interactions three and more, we can notice a slight increase in the number of frequency, which may mean that a larger number of previous visits from other sources results in an increase in brand awareness so that the customer starts visiting the website only through the *Direct Traffic* source.

Based on the analysis of Criterions 2 and 3, it can be concluded that according to the Li and Kannan (2014) study, there is a strong spillover effect of transmitting website visits

	Relative frequency			
No. of interactions	Company 1 (%)	Company 2 (%)	Company 3 (%)	Company 4 (%)
1	27.18	31.34	33.76	32.18
2	18.24	19.64	19.78	19.90
3	11.83	13.63	13.12	12.40
4	7.02	9.67	8.82	8.39
5	5.88	6.25	6.24	5.96
6	4.58	4.36	3.66	4.22
7	2.98	2.92	3.66	3.32
8 and more	22.29	12.19	10.97	13.6

Table V. Relative frequency of the length of customer journeys (Criterion 1)

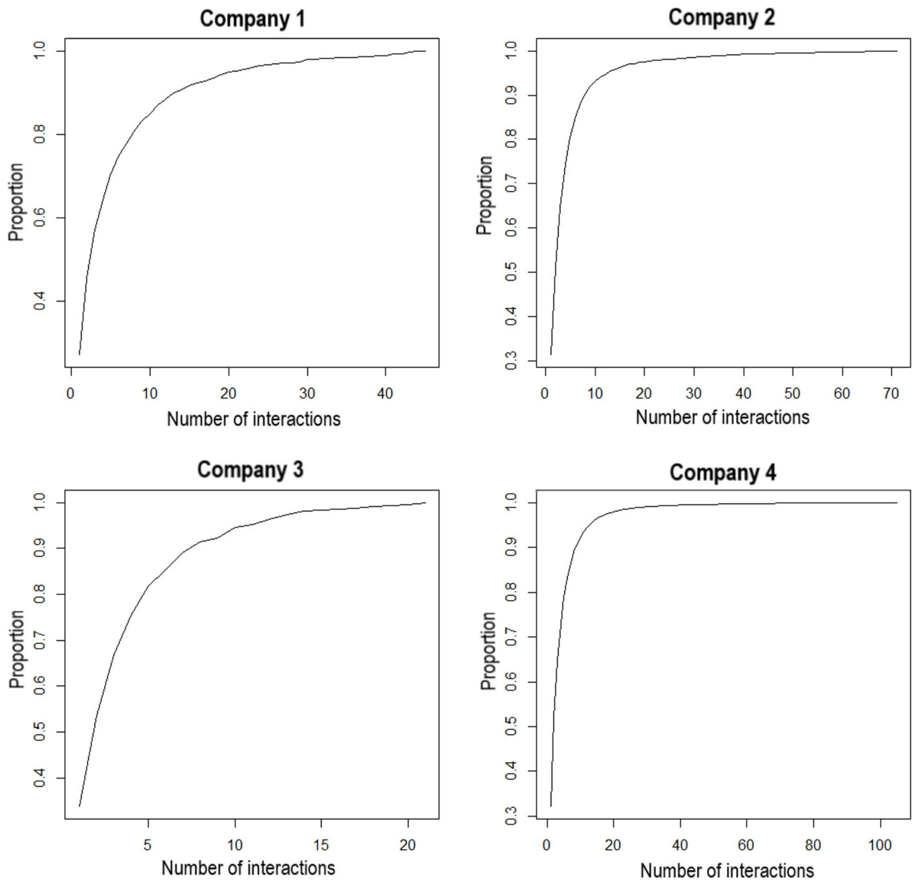

Figure 2.
Number of interactions – cumulative frequency (Criterion 1)

Table VI.
Relative frequency of interactions before reaching Criterion 2

No. of interactions	Relative frequency Company 1 (%)	Company 2 (%)	Company 3 (%)	Company 4 (%)
1	69.32	68.72	71.43	66.34
2	13.30	12.73	12.74	13.21
3	7.95	8.64	6.56	8.74
4	3.30	4.29	4.25	4.47
5	2.39	2.44	1.16	2.78
6 and more	3.75	3.17	3.86	4.46

from the previously used website traffic sources toward a specific source *Direct Traffic*, which results in the carryover effect when the customer further realizes his interaction with the website from the *Direct Traffic* source. The combination of these effects also has an impact on ending the customer journey in the form of a realized purchase.

At the beginning of this part of the work, a range of customers of the monitored companies, whose criteria for the *Direct Traffic* effect are related, was defined. However, its impact on company financial indicators – in this case, on generated revenue gained by the customers whose conversion journeys met the set criteria – also reflects the importance of the observed effect. The overview of the services generated on the basis of defined criteria is provided in Table VIII.

Table VII shows that for Companies 1, 2 and 3, the proportion of revenue in total sales is higher than the share of monitored customer journeys. This means that the value of these purchases is higher than the proportion of those purchases in total purchases. The purchases sorted out according to the set criteria are the most important for Company 1 because they account for about one-third of all purchases in their e-shop. Concerning Company 4, proportionately lower earning shares were recorded for purchases filtered based on Criterions 1 and 2. However, the absolute value of the sales of these purchases is clearly the highest among all the companies. Based on the monitored values, it can be said that *the Direct Traffic effect* deserves attention as a field of further research, as there is a possibility that optimization of brand awareness in the online environment can have a high added value for the company.

Conclusions and limitations

Customers' shopping decisions force marketing professionals to look at the customer journey beyond their last interaction before purchasing. The presented study aims to analyze the impact of brand awareness created online on customer's buying behavior, based on the current state of knowledge. Within the results of the interaction analysis of individual customer journeys focused on the *Direct Traffic* source, the term *Direct Traffic effect* was introduced. The source (marketing channel) *Direct Traffic* was placed on the level of the creator and the result of brand awareness in the online environment. During the customer journey, the customer moves from company-initiated interactions to interactions initiated by himself (Anderl *et al.*, 2015), which also results from the study carried out by

	Relative frequency			
No. of interactions	Company 1 (%)	Company 2 (%)	Company 3 (%)	Company 4 (%)
1	65.66	66.29	70.66	61.73
2	14.87	12.26	10.18	12.69
3	8.07	9.11	8.38	10.16
4	4.11	5.17	4.19	5.60
5	2.53	3.15	1.20	3.73
6 and more	4.75	4.03	5.39	6.09

Table VII. Relative frequency of interactions before reaching Criterion 3

	Company 1	Company 2	Company 3	Company 4
Total revenue	€1,579,778.00	€976,514.82	€219,719.49	€4,889,682.05
Criterion 1	€481,133.74	€260,403.89	€61,144.36	€1,295,538.59
Criterion 1 (%)	30.46	26.67	27.83	26.50
Criterion 2	€542,035.20	€180,311.11	€36,184.32	€947,862.87
Criterion 2 (%)	34.31	18.46	16.47	19.38
Criterion 3	€531,638.96	€128,507.20	€25,788.08	€686,486.70
Criterion 3 (%)	33.65	13.16	11.74	14.04

Table VIII. Revenues generated by defined criteria and their share in the total revenues

Li and Kannan (2014) concerning both *the carryover effect* and *the spillover effect*. There are three criteria that support the conclusions of the previous studies. It was found that each of the specified criteria (in absolute values) accounts for a significant share of the company's revenues. In the case of customer-initiated visits, a company does not pay for customer interaction. In conjunction with the theory of Barreda *et al.* (2015), the company can build awareness among prospective customers by offering them a good customer experience during the first interactions. This will brand on the customer's memory so that he will not only repeatedly visit the website from the *Direct Traffic* source but also his customer journey will end with the purchase of the company's products. Based on the *Direct Traffic effect* results, it is also possible to see that for more than 60 per cent of customers, only one previous interaction with the company's website is sufficient.

The results of this analysis can be influenced by factors that could not be taken into account during its implementation. The limitations are as follows:

- *Data collection at the cookie level*: as mentioned in the previous sections of the study, the user data might be considered to be data regarding the user's single device because of cookies. Thus, if a user uses more than one device, the Web analytics software [unless cross-device tracking is set (Alhlou *et al.*, 2016)] will record him multiple times as a different user. Additionally, if a user deletes cookies in a Web browser, the Web analytics software will record him as a new user. Cookies are, however, according to Flosi *et al.* (2013) standard for tracking in multichannel analytics;
- *Customer journeys lasted up to 30 days*: some customer journeys could last longer than 30 days. This could cause the first interaction of the actual customer journey that might occur in the past and were not recorded, which could have been the cause of distortion of the attribution modeling results. The *Direct Traffic effect* analysis could also filter customer journeys that started with other channels than *Direct Traffic*. Because these interactions took place earlier (as monitored 30 days long window), the analytics system could record the *Direct Traffic* source as the first source of the visit;
- *Customer journeys represented interactions with the website*: the customer could also come into contact with the marketing communications of companies elsewhere than on the company's website. For example, he could see an ad and not click on it, look through a page on social networks (like the Facebook page) and not click on a website, etc. Such behavior was not included in customer journeys; and
- *The ambiguity of the Direct Traffic source*: the *Direct Traffic* source could represent one of the other marketing resources, e.g. a visit from a mobile app (Facebook and Messenger), a browser bookmark or an offline ad such as billboards, leaflets or catalogs. In addition, each of the analyzed companies has a bricks-and-mortar shop. However, by our selection of the companies, we have tried to eliminate the impact of offline advertising.

Future research should focus on the elimination of the abovementioned limitations. Moreover, as the brand awareness, if not supported, decays over time (Binet and Carter, 2018), we would like to examine the potential difference in the contribution of *Direct traffic* into generated profit in various timing conditions. Additional research should also focus on finding the particular elements of the website that drive customers into remembering the brand/website name while visiting the website. These elements might affect the future online revenue of the companies, and therefore, by knowing its impact, companies are able to improve customer's experience in accordance with Lemon and Verhoef (2016).

References

Abhishek, V. Fader, P. and Hosanagar, K. (2012), "Media exposure through the funnel: a model of multi-stage attribution", available at: http://poseidon01.ssrn.com/delivery.php?ID=719105092024097125112107017080008118038069081083039091121092126090086107098067028066123029119022059121027020064094087095094066017078060077029075098065095004124065113085042032005064009082104092003022031091072098029023076127099123106002079010012087124078&EXT=pdf (accessed 24 September 2018).

Alhlou, F., Shiraz, A. and Fettman, E. (2016), *Google Analytics Breakthrough: From Zero to Business Impact*, John Wiley and Sons. Hoboken, p. 625. ISBN 978-1-119-14401-4.

Anderl, E., Schumann, H.J. and Kunz, H.W. (2015), "Helping firms reduce complexity in multichannel online data: a new taxonomy-based approach for customer journeys", *Journal of Retailing*, Vol. 92 No. 2, pp. 185-203.

Anderl, E., Wangenheim, F. and Schumann, H.J. (2016), "Mapping the customer journey: lessons learned from graph-based online attribution modeling", *International Journal of Research in Marketing*, Vol. 33, pp. 457-474.

Anderl, E. Becker, I. Wangenheim, F.V. and Schumann, J.H. (2014), "Mapping the customer journey: a graph-based framework for online attribution modeling", available at: http://ssrn.com/abstract=2343077 (accessed 24 September 2018).

Ash, T. (2012), *Landing Page Optimization: The Definitive Guide to Testing and Tuning for Conversions*, John Wiley and Sons. Indianapolis, ISBN 978-0-470-61012-1.

Barreda, A.A., Bilgiham, A., Nusair, K. and Okumus, F. (2015), "Generating brand awareness in online social networks", *Computers in Human Behavior*, Vol. 50, pp. 600-609.

Binet, L. and Carter, S. (2018), *How Not to Plan: 66 Ways Not to Screw It Up*, Matador. Leicestershire, ISBN 978-1789014-501.

Bones, C., Ham Mersley, J. and Shaw, N. (2019), "Optimizing digital strategy: how to make informed", *Tactical Decisions That Deliver Growth*, Kogan Page. London, ISBN 978-0-7494-8781-2.

Bryl, S. (2016), "Attribution model with R (part 1: Markov chains concept)", available at: http://analyzecore.com/2016/08/03/attribution-model-r-part-1/ (accessed 24 September 2018).

Clifton, B. (2015), *Successful Analytics: Gain Business Insights by Managing Google Analytics*, Advanced Web Metrics, West Sussex, p. 336. ISBN 978-1-910591-00-0.

Danaher, P.J. and Van Heerde, H.J. (2018), "Delusion in attribution: caveats in using attribution for multimedia budget al.location", *Journal of Marketing Research*, Vol. 1, pp. 1-65.

Dinucci, D. (1999), "Fragmented future", *Print*, Vol. 53 No. 4, p. 32, available at: http://tothepoint.com/fragmented_future.pdf

Ferencová, M., Jelenõová, I. and Kakalejčík, L. (2015), "Social media usage in product information searching", *Applied Mechanics and Materials*, Vol. 795, pp. 69-76.

Finstat (2017), Dostupné na Internete, https://finstat.sk/ (accessed 27 January 2018).

Geyik, S.C., Abhishek, S. and Dasdan, A. (2014), "Multi-touch attribution based buget al.location in online advertising", *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*: ACM. New York, NY, pp. 1-9, ISBN 978-1-4503-2999-6.

Ghose, A. and Todri, V. (2015), "Towards a digital attribution model: measuring the impact of display advertising on online consumer behavior", *MIS Quarterly*, Vol. 1, ISSN 2162-9730.

Google Analytics (2018), Dostupné na Internete: https://analytics.google.com/ (accessed 27 January 2018).

Kahn, B.E. (2013), *Global Brand Power: Leveraging Branding for Long-Term Growth*, Wharton Digital Press. Philadelphia, p. 134. ISBN 978-1613630266.

Kakalejčík, L., *et al.* (2018), "Multichannel marketing attribution using Markov chains", in: *Journal of Applied Management and Investments*, Vol. 7 No. 1, pp. 49-60.

Karlin, S. and Taylor, H. (1975), *A First Course in Stochastic Processes*, Academic Press, Cambridge.

Kim, H.K., *et al.* (2008), "Brand equity in hospital marketing", *Journal of Business Research*, Vol. 51 No. 1, pp. 75-82.

Lemon, K.N. and Verhoef, P.C. (2016), "Understanding customer experience throughout the customer journey", *Journal of Marketing*, Vol. 80 No. 6, pp. 69-96.

Li, H. and Kannan, P.K. (2014), "Attributing conversions in a multichannel online marketing environment: an empirical model and a field experiment", *Journal of Marketing Research*, Vol. 51 No. 1, pp. 40-56.

Marshall, P. and Todd, B. (2017), *Ultimate Guide to Google AdWords: How to Access 100 Million People in 10 Minutes*, Entrepreneur Press, Irvine, p. 380. ISBN 978-159918612-2.

Nissar, T. and Yeung, M. (2015), "Purchase conversions and attribution modeling in online advertising: an empirical investigation", EMAC 2015, Leuven.

Rentola, O. (2014), *Analyses of Online Advertising Performance Using Attribution Modeling: diplomová Práca*, Aalto University, Aalto, p. 71.

Roberge, M. (2015), *The Sales Acceleration Formula: Using Data, Technology, and Inbound Selling to Go from $0 to $100 Million*, John Wiley and Sons, Hoboken, p. 224. ISBN 978-1-119-04707-0.

Romaniuk, J. and Sharp, B. (2015), *How Brands Grow: Part 2: Emerging Markets, Services, Durables, New and Luxury Brands*, Oxford University Press, Oxford: ISBN 9780195596267.

Shao, X. and Li, L. (2011), "Data-driven multi-touch attribution models", *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM New York, NY, pp. 258-264, ISBN 978-1-4503-0813-7.

The R Project for Statistical Computing (2016), "The R project for statistical computing", available at: www.r-project.org/. (accessed 19 August 2016)

Van Den Berg, G. and Pietersma, P. (2015), *25 Need to Know Management Models*, Pearson. Harlow, p. 224. ISBN 978-1-292-01635-1.

Vanden Heuvel, A. (2014), "5 Ways to nurture leads through the sales process, eCommerce style", Hubspot, 15 April, updated February 01 2017, available at: https://blog.hubspot.com/insiders/lead-nurturing-tips

Wang, Q., Li, B., Wang, P. and Yang, J. (2015), "When structural models meets big data: examining multi-device attribution for native ads using TB-sized data", *Proceedings of Workshop on Information Systems and Economics*, available at: https://research.yahoo.com/mobstor/publication_attachments/LargeScaleStructualAnalysis.pdf

Wooff, D.A. and Anderson, J.M. (2015), "Time-weighted multi-touch attribution and channel relevance in the customer journey to online purchase", *Journal of Statistical Theory and Practice*, Vol. 9 No. 2, pp. 227-249.

Yadagiri, M.M., Saini, S.K. and Sinha, R. (2015), "A non-parametric approach to the multi-channel attribution problem", *16th International Conference*, Springer International Publishing, Basel.

Corresponding author
Jozef Bucko can be contacted at: jozef.bucko@tuke.sk

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**