Guided by the crowd: herding in Untappd ratings

Silvan B.P. Mensink

940802558020

April 2020

MCB-70424

Supervisor:

Dr. Nikita Sharda

Marketing and Consumer Behaviour Group

Second reader:

Dr.ir. Frans J.H.M. Verhees

Marketing and Consumer Behaviour Group

WAGENINGEN UNIVERSITY & RESEARCH

Contents

Abstract

The rise of numerous online community platforms that are easily accessed to gather and share information, made that we increasingly rely on others' aggregated online opinions when making decisions for ourselves. Especially in growing industries such as craft beer, consumers do not hesitate to share their opinions via their mobile device(s). Creating new opinions while relying on others', ensures that we engage in the act of herding. However, scholars have found mixed results when trying to determine the direction of herding on online community platforms. Therefore, it is important to investigate online community platforms that are dedicated to product categories such as craft beer. This research used online community platform Untappd as a natural environment to test for the direction of the herding effect in ratings for craft beer. It was found that the overall rating tendency of users on the platform is quite positive, and that subsequent user ratings are more positive than the aggregated rating. Moreover, unexpected results showed neutralisation effects in the extremities of the rating scale. The main implications are addressed to scholars, trying to model the effect of prior ratings on subsequent ratings, and managers in the craft beer industry, to take strategic marketing decisions.

1. Introduction

Craft beer consumption is booming (CGA Strategy, 2019; Modor Intelligence, 2019). More specifically, a growing number of consumers share their opinions about craft beers to others using social media platforms (Untappd, 2019). In 2000, RateBeer was one of the first online community platforms (forum) where consumers of craft beers could exchange opinions and rate what they drank. In January 2012, more than 3,500,000 beers were reviewed and rated. A decade after the launch of Ratebeer, "socially drinking" community platform Untappd was founded. The approach of Untappd was unique, as was focused for use on mobile devices. To date, Untappd is the largest craft beer community, with more than eight million users and 800 million user contributions (Untappd, n.d.; 2019).

As the slogan "drink socially" suggests, Untappd anticipates on the growing number of consumers that are willing to share opinions about consumption-related products with others (Brown et al., 2007). Untappd is a free mobile application (app) that can be download on mobile devices running iOS or Android operating systems (OS). iOS and Android combined make up for 99% of the mobile OS market share in the world (Statista, 2020). This ensures Untappd to reach almost all consumers with a mobile device. One that opens the app for the first time is forced to create a personal account, which can either be relatively anonymous via e-mail or more personal via the Facebook community platform. Once the account has been created, the user is encouraged to share craft beers, breweries and venues with others on the platform. This user created content is called a "check-in". When the user searches for a craft beer they are currently drinking, they are redirected to the corresponding page of the craft beer (Figure 1). On this page, information such as the style, alcohol percentage and the brewery are shown. Moreover, three types of ratings are displayed: 1) the average global rating, 2) the average friends rating, and 3) the users' own rating. Besides, a big green check-in button appears. Once the user taps the button, they are redirected to the check-in page (Figure 2), where they are able to write a small review (of a maximum of 140 characters) and rate it on a 0 to 5 point scale. Additionally, the user can upload a photograph, indicate in what kind of way the craft beer was served, tag friends and the flavour profile, and the venue where they are consuming the craft beer. Once the user is finished and taps the check-in button again, the checked-in craft beer is logged and displayed on the global activity feed (Figure 3), where others can comment or toast (i.e., similar to a Facebook "like") upon. More specifically, when a user becomes "friends" with other users by sending or accepting friend requests in the app, checked-in craft beers of the user itself and its friends are prioritised in the "activity" feed.

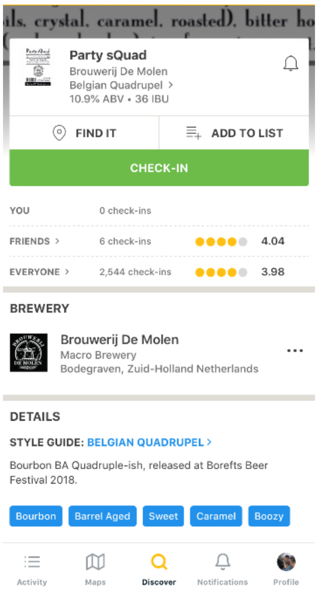


Figure 1: Craft beer page

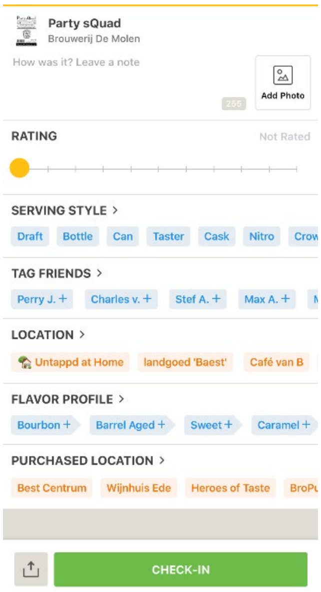


Figure 2: Check-in page

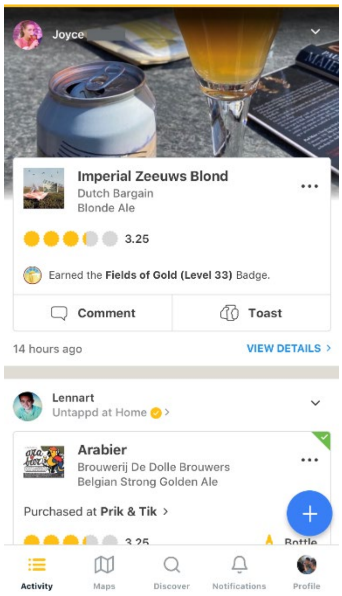


Figure 3: Activity feed

In sum, Untappd has a great social character which is characterised by three main components in the app. The first being the page dedicated to the craft beer a user searched, showing three types of user ratings. Second, the check-in page, where the user is able to judge the craft beer for themselves and show their judgement to others using the app. The check-in page can only be reached through the page dedicated to the craft beer. Third, the activity feed, where own and friend activity is shown. The latter can be reached independently of the other pages.

The way a user is directed through the Untappd app to perform a check-in, assures that they are exposed to one or more rating types before they reach the check-in page. That is, a user is unable to perform a check-in without begin exposed to the average global rating, friend rating and/or own rating. A new check-in of an user exposed to one or more rating types is thus inevitably influenced by prior ratings of others (Tucker and Zhang, 2011), and does therefore not necessarily reflect the true quality of the product (Hu et al., 2009; Gao et al., 2015). This claim is far from unique, as it has been confirmed in multiple studies and is referred to as herding effect or social influence bias (e.g. Schlosser, 2005; Li and Hitt, 2008; Godes and Silva, 2012; Tucker and Zhang, 2011; Guo and Zhou, 2016). However, researchers are unable to determine the direction of such a herding effect in independent online community platforms. That is, inconsistent empirical findings were reported about whether there is a positive or negative relationship of prior ratings on subsequent ratings (Ma et al., 2013; Wang et al., 2018; Lederrey and West, 2018). Moreover, it is important to determine the direction of the herding effect in the craft beer industry, as consumers rely on reviews and ratings when deciding what products to buy (Zhu and Zhang, 2010; Moe and Trusov, 2011; Godes and Silva, 2012). Positive herding could be beneficial for craft beer companies, as it could account for a more positive rating, eventually leading to more sales. However, negative herding could be damaging, as it could lead to a more negative rating and less sales. This paper is the first to shed light on the direction of the herding effect in Untappd ratings specifically, and provides some strategic recommendations, in such that it will be particularly useful for the increasing number of craft beer companies in the world. Therefore, the main question is:

What is the direction of the herding effect in Untappd ratings?

To answer this question, the following sub-research questions need to be answered:

1. What are antecedents for a herding effect in ratings in online community platforms?
2. How is a herding effect in ratings in online community platforms formed?
3. Is there a herding effect present in prior ratings on subsequent ones on Untappd?

It is worth noting that Muchnik et al. (2013) found differences in the extremes of the rating scale, where positive and negative aggregated ratings got *both* more positive over time. Positive ratings were accumulated, negative ratings were neutralized. Muchnik et al. (2013) argued that more empirical evidence is needed, to investigate whether the effect they found is also present in other industries. As this is in line with this research, the second main question arose, assuming that the herding effect is present:

In line with Muchnik et al. (2013), are there differences found in the extremities of the rating scale?

To answer this question, the following sub-research questions need to be answered:

1. How far is the subsequent rating away from prior ratings in the upper extreme?
2. How far is the subsequent rating away from prior ratings in the lower extreme?

This paper is built up as follows: First, a literature background is given, where the antecedents and the formation of the herding effects in online community ratings is discussed. Second, the research method is discussed. Third, using web scraper data, the existence and direction of the herding effect is tested in ratings of the Untappd community platform. Additionally, using web scraper data, extreme ratings are outlined, to test the theory of Muchnik et al. (2013). Lastly, results, conclusions, limitations and implications are discussed as well as recommendations for future research.

2. Literature background

In this part, literature regarding herding effects in rating behaviour on online community platforms is discussed. More specifically, literature in the realm of word of mouth (WOM) and online or electronic word of mouth (eWOM), in which herding effects get a lot of attention by marketing researchers. Key to this part is to define antecedents and the formation of such a herding effect. By using this key information, hypotheses are formed to be applicated to answer the main research questions.

2.1. Antecedents of herding effects in online communities: WOM and eWOM

In the past decades, extensive literature focused on and showed the importance of (offline) WOM on consumer's attitudes and purchase behaviour (e.g. Brown and Reingen, 1987; Richins and Root-Shaffer, 1988; Herr et al., 1991). WOM is unpaid consumer-to-consumer interaction or communication (Reingen and Kernan, 1986; Sun et al., 2006). WOM has great influential power, as it usually stems from someone within our personal circle (Meuter et al., 2013). This makes the communicated recommendation (which is interpersonal in nature) to be experienced more authentic and reliable by the receiver than other forms of communication, such as external communication like advertisements (Duhan et al., 1997). More recently, eWOM, or online WOM, emerged as it became more readily available due to the numerous and increasing amount of online tools and social media apps (Meuter et al., 2013), also described as the Web 2.0 (Jansen et al., 2009; O'Reilly, 2005). Zhu and Zhang (2010) found empirical evidence that online consumer reviews were a good proxy for traditional offline WOM. Moreover, they found that online consumer reviews – just like WOM – influenced consumers' decisions, which eventually influence sales. Drawing on the definition of WOM, Hennig-Thurau et al. (2004) defined eWOM, as "*any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet*" (p. 39). This definition is mostly used across journals and papers. Hennig-Thurau et al. (2004) found that most eWOM happens trough web-based opinion platforms, and that, unlike WOM, eWOM expert and consumer opinions are easily accessible via the internet. Furthermore, they provide consumers the opportunity to read reviews and ratings of a product, service or company long after the reviewer has posted their opinion.

Given the abovementioned findings, it is of no surprise that eWOM gained popularity in past and present electronic commerce (ecommerce) (e.g. Yin et al., 2014) and social network (e.g.

Guo and Zhou, 2016) research. Nowadays, many consumers infer product quality and match of preference from product reviews and ratings on the internet (Kwark et al., 2014). At the beginning of this century, major companies like Amazon and Apple, but also small and local ones have implemented rating mechanisms in their purchasing platforms to benefit from it as a new marketing tool (Dellarocas, 2003). Moreover, independent online community platforms (i.e. Facebook and Untappd.com) emerged where consumers are able to create and read product reviews and ratings. These independent online community platforms have shown to be more influential than company-controlled sources (Meuter et al., 2013; Brown et al., 2007; Smith et al., 2005). Additionally, researchers showed that due to the rise of Web 2.0 and social media (Lai and Truben, 2008), and the popularity of mobile devices and apps (Thakur, 2016), consumers engaged even more into consumer-to-consumer and consumer-to-company interactions. They do this by for example creating, reading and sharing reviews and ratings for products and services before or after they purchase them (Shankar et al., 2010; Hajli, 2014).

The rise of these online communities and the ease to access them, ensured that more specified communities emerged, in which individuals with an interest in a particular product category participated (Brown et al., 2007). Most frequently, these communities are consumption-related, where one is able to show or express their knowledge and preference. Through a single eWOM expression of an individual in these online communities, one message could easily reach and influence many other consumers using the community platform (Lau and Ng, 2001; Park and Lee, 2008). Moreover, it is assumed in social network research that subsequent decision making of an individual who is involved in a particular online community is influenced by eWOM (Brown et al., 2007). This could be problematic, as the first eWOM expressions about a product are often created users that are opting-in (i.e., intrinsically motivate to give a review or rating), causing two types of selection-biases (Hu et al., 2009): First, purchasing bias occurs because of people that are motivated to buy a certain product will most likely be reporting positively about it. Second, underreporting-bias takes place as people that are extremely positive of extremely negative about the product are more likely to express their opinion compared to people with moderate opinions. It is thus found that early consumers of a product set the default rating of the product for the entire online environment, accessible to everyone (Li and Hitt, 2008).

Given these findings, the strength eWOM is significant in online communities, paving the way for others being influenced by eWOM expressions of others within these online communities. This phenomenon is called social influence bias or the herding effect, where a subsequent user opinion within a community is influenced by a prior one. In the next section, the formation of herding effects in online community platforms is discussed.

2.2. Formation of herding effects in online community platforms

According to ecommerce research, there are two possible routes for one to be influenced by others trough eWOM: the informational and normative route (Park and Lee, 2008). One takes informational influence route when they accept the information about a product given by others on the online review platform. The normative route is taken when one conforms to the group expectations. That is, one values the product equal to other users of the online review platform. Informational and normative routes can be taken simultaneously as sources of information to consumers: product information (informational route) as a qualitative information source and recommendation information (normative route) as a quantitative information source. It is worth noting that high-involvement users are more likely to take the informational route, as they are particularly interested in focal and extensive information about a product (Park and Lee, 2008). Low-involvement users rely more on recommendation information, that is, peripheral cues such as simple ratings of a product. Overall, both routes indicate whether a product is being liked or disliked by users of the online review platform, and influence others to value the product the same way, regardless of their involvement with the product (Park and Lee, 2008; Venkatesan, 1966).

More specifically in social network research, it is assumed that subsequent decision making of an individual who is involved in a particular online community is influenced by eWOM in terms of tie strength, homophily and source credibility (Brown et al., 2007). Brown et al. (2007) found that the greater the tie strength, homophily and source credibility, the greater the chance of a herding effect. As all WOM takes place within the personal circle, the strength of the closeness or relationship between source and receiver of the information is defined as tie strength (Money et al., 1998). A strong tie strength means more WOM interactions and has a greater influence on consumer behaviour compared to weaker tie strength. This is in line with Schlosser (2005), who found that people post more positive messages if that is favourable to the relationship between sender and receiver(s), regardless of the actual experience with the product they are rating. This also applies the other way around. Additionally, Esch et al. (2006) argued that (intimate) relationships do not necessarily need to be between two people but can also exist between people and brands. Brown et al. (2007) extended this statement and found that this relationship also exists between people and the online community platform. Ultimately, a strong relationship between the person and the online community platform would influence the persons' behaviour to the opinion of the community platform.

Related to tie strength, the construct homophily encompasses the similarity of online community members' characteristics, such as age, gender and lifestyle (Rogers, 1983), and also in terms of shared interest and similarity of situation (Schacter, 1959). Therefore, the more the two individuals are alike (homophile), the greater their tie strength is (Brown et al., 2007; Wang et al., 2018). However, as online communities are basically anonymous due to the nature of the internet, it is almost impossible to infer homophily from social characteristics. Therefore, other comparisons of similarity, such as inclusive mind-set or shared group interests, come into play (Brown et al., 2007). Inclusive mind-set in online communities is expressed by showing gratitude to other users. Shared group interests shown by a user responding that they feel the same way. For example, one thanks another user for sharing a product rating or responds to a message by saying that they also liked the product.

Lastly, source credibility is an important construct, as online community members infer the competence of the source from the senders' expertise and online community platforms' trustworthiness. As an example, perceived credibility does not only depend on the contributors' review or rating of a product but is also influenced by the online community platform itself (Brown et al., 2007).

In sum, ecommerce and social network researchers report a direct effect of prior reviews on subsequent ones. For both paradigms, it shows the presence of herding effects in ratings. In ecommerce, regardless of the involvement and the route that is taken, consumers' decisions are influenced by prior ratings (Park and Lee, 2008). Moreover, in social network research, tie strength, homophily and source credibility in online community platforms through eWOM are important constructs that influence consumer decision making (Brown et al., 2007). These findings inevitably pave the way for herding effects within these online communities. Brown et al. (2007) showed that individuals are not only influenced by other individuals using these online communities, but also the online community themselves through tie strength, homophily and credibility. This is a major issue of online communities, as it means that the content that is submitted to these platforms (e.g. product reviews and ratings) are not self-contained but influenced by others or platforms. That is, subsequent reviews or ratings are conditional to prior reviews and ratings (Moe and Schweidel, 2012). Tie strength, homophily and source credibility of users of a platform and the platform itself influence other users' ratings, regardless on their involvement to the product category. The direction of the effect, however, is dependent on a variety of reasons. As this paper is focused on the direction of the herding effect in Untappd ratings, only factors that are present in the Untappd platform are considered in the hypotheses in the next section.

2.3. Hypotheses development

In this section, hypotheses are developed. Given the functions of Untappd that are explained in the introduction section and the antecedents and formation of the herding effects in online community platforms, Untappd and online community platforms can be compared to each other. By comparing Untappd with other online community platforms, the assumed direction of the herding effect in Untappd ratings is substantiated. Moreover, regarding the second research question, the direction of the ratings in the extremes are predicted using the functions of Untappd and the literature.

2.3.1. Direction of prior ratings on subsequent ratings in Untappd

The Untappd platform can be categorised as an online recommender system for craft beer (Goswami and Kumar, 2019). The online community platform provides its users with short and perceivably helpful information during the evaluation (reviewing and rating) process. Given that a herding effect is present on independent online social product rating platforms such as Untappd (e.g. Ma et al., 2013; Wang et al., 2018; Lederrey and West, 2018), and that herding effects are more prominent in the evaluations of products such as craft beer (Li and Wu, 2018), it is most certainly that there is a herding effect in Untappd ratings. The direction however should be determined on basis of the findings of the online community platforms that have been researched. Ma et al. (2013) investigated an online community platform for rating (local) businesses and Wang et al. (2018) online books.

Ma et al. (2013) found that prior experience, geographic mobility, social connectedness, gender, length of the review and time interval are significantly moderating the relationship between prior ratings and subsequent ones. They posited that less experienced and mobile people are more prone to herding than more experienced and geographically mobile people. Moreover, they mention social connectedness (which is a similar construct to tie strength (Brown et al., 2007)) positively moderates the relationship as reviewers adhere to the normal. Furthermore, gender is mentioned as a moderating factor: males are more prone to prior reviews as they try to detect the general idea of others about products, before making their own decision. According to Ma et al. (2013), women take a more informal route and are therefore less vulnerable to other's opinions. Additionally, Ma et al. (2013) found that review characteristics (length of the review and time interval between reviews) significantly influence the relationship between prior ratings and subsequent ratings. Longer reviews reduce the likelihood of a subsequent review to be influenced by a prior one, which is consistent with the findings of Park and Lee (2008). As longer reviews are more qualitative in nature, people weigh the arguments and subsequently form their own opinion (Park and Lee, 2008). Lastly, Ma et al. (2013) found that longer time intervals between a users' reviews positively moderates the relationship between prior ratings and subsequent ones.

Wang et al. (2018) found that people rate in a similar way when friendships on the online community platform are formed. This is finding is similar to the concepts of tie strength and homophily, as described by Brown et al. (2007). Interestingly, people with fewer friends on the platform rate more easily influenced by others as well. In line with Moe and Schweidel (2012), Wang et al. (2018) found that more experienced users and heavily rated books negatively influenced the relationship. Moreover, they found that after controlling for book characteristics that could influence the relation between prior and subsequent ratings the relationship was negative. They attribute this to the rater's tendency to diverge from public ratings (Moe and Trusov, 2011).

As Untappd is an app dedicated to social craft beer consumption, its users are most likely to be interested in the product category. According to Chorley et al. (2016), the majority of Untappd users (65.2%) check-in occasionally (less than once per 10 days). However, a small group of users (2.4%) checked-in more than one craft beer a day. Most users of Untappd are thus less experienced with all kinds of different craft beers, which positively affects the herding direction in Untappd ratings. Ma et al. (2013) posited that geographical mobility would influence the relationship of prior ratings on subsequent ones. Although, this may be a great issue when researching restaurants, a craft beer can easily be ordered a long time prior to consumption. It is thus not seen as an issue in the case of Untappd. Social connectedness (Ma el al., 2013) and friendships (Wang et al., 2018) are very interesting concepts, also introduced by Brown et al. (2007) as tie strength and related to homophily. Chorley et al. (2016) found that Untappd users have an average friendship group size of 9 on the platform. Although this may sound very little, they showed that users that are friends have a similar taste for types of craft beer. This indicates that craft beers that have been drunk by a particular user also have a great chance to be drunk by his friends. This affects the herding direction positively. Ma et al. (2013) found that males are more vulnerable for herding than women. It is no secret that there are more men drinking alcoholic beverages than women. Wilsnack et al. (2009) showed in a research conducted worldwide that in 98 out of the 104 cases, men drank more than women. It is not specifically known what the gender ratio is for the Untappd platform. The best guess is thus that, because there are more males drinking alcoholic beverages, there are more male Untappd users than female users. Therefore, it is more likely that users engage in the act of herding, given that positive herding is more prevalent with males than females. Wang et al. (2018) found that rating volume would influence the direction of the herding effect. However, in a paper that specifically investigated Untappd data, this was not the case (Chorley et al., 2016).

Comparing other online community platforms to Untappd in terms of the factors that would determine the direction of the herding effect on Untappd, it is assumed that they count for a positive herding effect. However, the effect should be tested for the Untappd platform specifically. Recall that a user is insurmountably exposed to one or more rating types in the Untappd app, and that "average global rating" is the only rating type that is always determined and visible (i.e., it is possible that none of the user's friends or that the users themselves checked-in). It is therefore stated that:

H1: The average global rating positively influences subsequent ratings of users on the Untappd platform.

2.3.2. Direction of prior ratings on subsequent ratings in Untappd in the extremities

As stated in the introduction, Muchnik et al. (2013) found differences in the extremities of the rating scale, indicating an asymmetric herding effect. In their study, they manipulated newly submitted social news posts, by either treating them down, up or not at all. They showed that herding in extreme positive ratings accumulates, leading to more positive subsequent ratings. Moreover, extreme negative ratings were neutralised by crowd correction, meaning that they tended to get more positive over time. They attributed the accumulating positive rating trend to people with less rating experience that engaged in rating posts. The neutralisation of negative ratings was attributed to the tendency of users to rate positive on the site that was investigated.

In another study specifically for beer ratings, it was found that newly submitted beer ratings that received a very high first rating, also received a high subsequent rating (Lederrey and West, 2018). This trend also applied to beers that received a very low first rating: the subsequent rating was also low. The effect was still noticeable after more than twenty ratings were given for that particular beer. Lederrey and West (2018) did not find evidence for a neutralisation effect, as Muchnik et al. (2013) did. They attribute their findings purely to herding behaviour of users using the platforms that they investigated (BeerAdvocate and RateBeer).

Given the abovementioned findings in the extremities of the rating scales, it is assumed that Untappd users' herding behaviour is closer to the findings by Lederrey and West (2018), as they researched platforms that are around a similar product category. Moreover, it is not known what the rating tendency of Untappd users is at the moment, which was given as a reason by Muchnik et al. (2013) for the neutralisation effect in extreme negative news post ratings. It is therefore stated that:

H2a: Extreme positive average global ratings positively influence subsequent user ratings on Untappd

H2b: Extreme negative average global ratings negatively influence subsequent user ratings on Untappd.

Although conducted research already gave an impression on the direction of the herding effect in general (Ma et al., 2013; Wang et al., 2018; Muchnik et al., 2013), and specifically for beer (Lederrey and West, 2018), they did not do this for ratings on the Untappd platform. Moreover, they took different approaches, as Lederrey and West (2018) only researched a possible herding effect in beers that were newly introduced to the platform. Additionally, all researched online community platforms were accessed by users via desktop- or laptop-ish devices, and not via mobile devices (i.e. tablets or smart phones). All factors combined make this research unique. The research method however can be derived from prior papers that researched the relationship between prior and subsequent ratings. The methodology is explained in the next section.

3. Research methodology

In this section, the research methodology is explained to show the direction of the herding effect between prior and subsequent ratings on the Untappd platform. A great tool to test this effect is a web crawler or web scraper, which is typically used in social platform analysis (Glenski and Weninger, 2017). Therefore, first, the web scraper tool is explained. Second, it is elaborated on the sample and conducted test.

3.1. Data collection

As a means to collect data, a web scraping tool is used. A web scraping tool is an automated data collection agent that enables one to research (commercial) web sites (Glenski and Weninger, 2017). The tool enables web data collection without manual interference. This ensures that data is systematically captured within a very short time frame. Today's web scrapers are able to identify certain "blocks" or "elements" on web sites (Debnath et al., 2005). These elements are web site components based on their content (i.e. images, texts, counters, etc.). Researchers are able to program the web scraper tool to systematically scrape a certain element (e.g. the rating of a certain person for a certain craft beer). By using this method, a lot of ratings can be captured systematically in a short amount of time. However, it is limited to the web site's boundaries (Glenski and Weninger, 2017). This means for example that programmers of the web site decide how many records are shown in total.

As Untappd is a platform that is focused on mobile devices, it may seem hard to extract data using a web scraper tool, simply because web sites and apps do not necessarily use the same technologies. However, Untappd does have a web site, which offers its users the functionality to view other's check-ins. Users are not able to check-in craft beers via the web site. This functionality is reserved for app users. A web scraping tool can thus still be used using the Untappd web site, as it displays all data generated by users.

The web scraping tool used for this research is webscraper.io (https://www.webscraper.io). Webscraper.io is fairly easy configured by pointing and clicking on elements on web sites and requires no coding. Moreover, it is free to use. Webscraper.io generates raw data, which can be exported to tools used for statistics. This makes the tool very interesting and applicable for this research.

3.2. Sample and conducted test

As this research tries to identify the direction of a herding effect in the Untappd platform, and to test the herding direction in the extremes of the rating scale, it is chosen to divide the ratings into three distinct categories: "low", "middle", "high". Untappd uses a standardised rating scale ranging from 0 – 5. 0 means that one disliked the craft beer, 5 means that one really liked the craft beer. Chorley et al. (2016) found that the vast majority (64%) of all Untappd ratings are falling between 3 to 4. Fewer ratings are lower than 3 and higher than 4. Therefore, for average global rating or $AvgGlobalRating_{ij}$, where *i* is the user's rating given for craft beer *j*, category "low" considers ratings lower than 3 ($AvgGlobalRating_{ij} < 3$), category "middle" considers ratings between 3 and 4 ($3 \geq AvgGlobalRating_{ij} \leq 4$), category "high" considers ratings higher than 4 ($AvgGlobalRating_{ij} > 4$). $AvgGlobalRating_{ij}$ is compared to the individual user ratings $Rating_{ij}$ using a paired samples t-test, as one is already exposed to the $AvgGlobalRating_{ij}$ when they perform a check-in.

3.2.1. Sample

There are loads of breweries present on Untappd and even more craft beers. Through webscraper.io, it is not possible to take all craft beers into consideration. Moreover, although it is fairly easy to collect data, it would still cost a lot of money, computer power and time to track all craft beers on the platform. Furthermore, it is not known how many craft beers there are. Therefore, it is considered to monitor 50 craft beers in each category, giving a total of 150 samples. For each sample, 296 user ratings can be captured. This is the maximum amount of user ratings each craft beer on Untappd can generate on a page. In total, 44400 (150 * 296) ratings are captured. The craft beers in each category were randomly selected as follows: on 24-03-2020, webscrape.io was set to crawl `/thepub/`[1] every 5 minutes between 19:00 and 22:00. Each time, 25 craft beers were captured, with a total of 900 (180 / 5 * 25). All captured craft beers were loaded into an Excel file and were assigned a number listwise. Using Excel, a craft beer was selected randomly at a time. This was done in Excel, by first using the function `RANDBETWEEN(1;900)`. This function generated random numbers between 1 and 900, the amount of captured check-ins. However, these numbers contained duplicates, which is something that is not wanted. Therefore, function `RANK.EQ()` was used combination with

[1] Can be accessed under https://www.untappd.com/thepub/. This web page captures all new check-ins and displays them in real time.

COUNT.IF()-1. Using these formulas in combination with each other in column B, it was made sure that the list contained all numbers from 1 to 900 without any duplicates.

In sum, starting in column A, for each row r : RANK.EQ(A$_r$;AFIRST$_r$:ALAST$_r$)+COUNTIF($A$$_r$:A$_r$;A$_r$)-1. In order to stop the numbers from changing, columns A and B were fixed by copying and pasting them as fixed numbers. The exhaustive list of randomly selected numbers can be found in "7.1. List of randomised numbers". Using the randomly selected numbers, each category ("high", "middle", "low") was filled with a cap of 50. This was repeatedly done until all categories were filled with 50 craft beers.

3.2.2. Conducted test

The test that is conducted is based on Wang et al. (2018). Their method is closest to the approach this research is going to investigate: Wang et al. (2018) looked at the subsequent rating, based on the prior average friend rating. This research is looking at a subsequent rating, based on the prior average global rating. It is thus clear that both approaches are very much alike. The method was purely based on observational data and required no interference of the researchers. Recall that either case, at the time a user posts a new check-in on the Untappd platform, they have already been exposed to the average global rating. Therefore, it can be concluded that one that tries to check-in a new craft beer, is inevitably influenced (Tucker and Zhang, 2011).

To test the direction of herding effect in rating behaviour and in the extremities, the following tests were conducted: $Rating_{ij}$ were compared to $AvgGlobalRating_{ij}$ to calculate the direction of subsequent ratings on prior ratings using a paired samples t-test. Thereafter, the same was done for the "high" and "low" categories separately.

4. Results

In this chapter, the results of the conducted tests were presented. First, results regarding the data sample were shown. Second, it was explained how missing values and outliers were treated. Lastly, results regarding the hypotheses were presented.

4.1. Data sample results

Ideally, it would have been the case that 44,400 individual cases were captured. However, and not surprisingly, not all craft beers met the amount of ratings (296) that webscraper.io is possible to scrape. At the end of scraping 150 craft beers in total (50 craft beers in each rating category), it was ended up with a dataset with a total of 37,080 cases. This was eight percent less. However, a significant amount of ratings were still captured.

Similar to Chorley et al. (2016), it was found that the vast majority of the ratings were between 3 and 4. Chorley et al. (2016) reported 64% of their ratings captured in this category. In this sample which reported all average global ratings subsequently until all rating categories (“high”, “middle”, “low”) were filled, 75.4% of the ratings fell in this category (Figure 4). Less craft beers got a rating higher than 4 (16.5%) and very little a rating lower than 3 (8.2%).

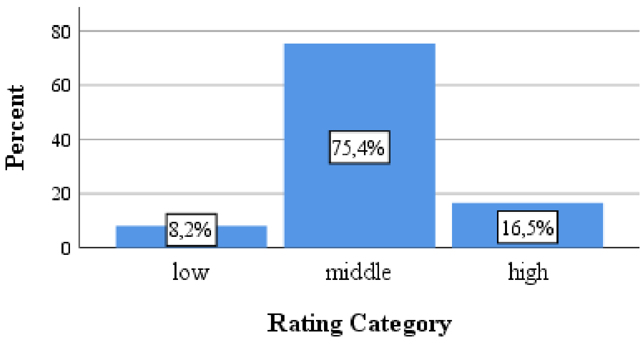


Figure 4: Percentage of ratings by Rating Category

4.2. Missing values

Of all cases captured, 6.1% missed a user rating. Three cases contained no information at all and were eliminated. Other cases were valid check-ins but did contain a subsequent rating (see “7.2. Missing values in dataset”).

In order to check whether those missing values are at random, cases are put into categories (case contains a missing value (0) or does not (1)). By running an independent samples t-test, the results show that there is no significant relationship between missing data in Rating_{ij} and the $\text{AvgGlobalRating}_{ij}$, $F(1,37075) = 2.516$, $p = .113$ (see “7.3. Check for relationship”).

Since there is no relationship found between $Rating_{ij}$ and $AvgGlobalRating_{ij}$, missing data is replaced by the series mean (SMEAN). SMEAN can be used only when it is assumed that there is no big difference between the mean of the $Rating_{ij}$ before and after treatment of the missing values. Therefore, a paired samples t-test is used to check the difference between the $Rating_{ij}$ before and after treatment.

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	$Rating_{ij}$	3.4704[a]	34824	.88829	.00476
	SMEAN($Rating_{ij}$)	3.4704[a]	34824	.88829	.00476

a. The correlation and t cannot be computed because the standard error of the difference is 0.

Table 1: Comparison of the standard error of the difference for $Rating_{ij}$ and SMEAN($Rating_{ij}$)

Table 1 shows that the standard error of the difference is 0, meaning that both variables $Rating_{ij}$ and SMEAN($Rating_{ij}$) are very alike regarding their mean. This indicates that the treatment of the missing values in $Rating_{ij}$ was successful. In the end, 37,077 cases were considered as valid.

4.3. Outliers

After treating the missing values, it was checked whether there were some outlier values that would skew the results. Cook's Distance (Cook's D) can be used to identify outlier values in cases (Field, 2013). Typically, values that are lower than 1 on Cook's D are of concern. However, it is also argued that this is more dependent on the size of the dataset, by setting a maximum Cook's D value of 4 divided by N. The latter is done by SPSS as a standard, and thus flags outliers when their Cook's D is > .00011 (4 / 37,077). To illustrate the outlier values, a scatter was plotted where the values of Cook's D by Centred Leverage Value are shown per case (see "7.4. Scatter of Cook's Distance by Centered Leverage Value").

The exclusion of outlier values left the dataset with 34,674 valid cases.

4.4. Hypotheses testing

After dealing with missing values and outliers, the model fit was first tested, to see how much of the variance in $Rating_{ij}$ is explained by the model (i.e. $AvgGlobalRating_{ij}$). Interestingly, the value that indicates the model fit (R^2) is .539, meaning that a significant amount of 53.9% of the variance in $Rating_{ij}$ is explained by $AvgGlobalRating_{ij}$, $R^2 = .539$, $F(1,34672) = 40478.61$, $p < .001$. This shows that $AvgGlobalRating_{ij}$ significantly predicts $Rating_{ij}$ which strengthens the assumption of herding in Untappd ratings. For output see "7.5. Regression output after treatment".

Subsequently, three paired samples t-tests were conducted for H1, H2a and H2b. Recall that a maximum of 296 ratings for a craft beer could be scraped from the Untappd we site. Since there are craft beers in the sample that have had exactly or less than 296 ratings, the average global rating would be determined solely by these ratings. To check whether there is a difference in the direction of ratings due to the low amount of ratings, the three hypotheses were also checked for craft beers with a minimum rating amount of 1000.

Hypothesis	Added requirement	Test indicator
H1		1a
H1	Total Ratings ≥ 1000	1b
H2a		2a
H2a	Total Ratings ≥ 1000	2b
H3a		3a
H3b	Total Ratings ≥ 1000	3b

Table 2: Indication of conducted tests

In the next three sections, the tests and its conclusions are presented (see Table 2). The tests are identified with a test indicator at the left side of the page. The test indicator represents the hypothesis and, if applicable, the added requirement.

4.4.1. H1: Average global rating on subsequent rating

1a

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SMEAN($Rating_{ij}$)	3.5424	34674	.7327	.0039
	$AvgGlobalRating_{ij}$	3.4876	34674	.6194	.0033

Pair 1 SMEAN($Rating_{ij}$) - $AvgGlobalRating_{ij}$

Paired Differences							
			90% Confidence Interval of the Difference				Sig.
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	(2-tailed)
.0549	.5043	.0027	.0504	.0593	20.255	34673	.000

A one-tailed paired samples t-test revealed that, in general, Untappd users gave a more **positive** subsequent rating (i.e., $Rating_{ij}$) compared to the average global rating (i.e., $AvgGlobalRating_{ij}$), $t(34673) = 20.255$, $p < .001$. H1 was accepted.

1b

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SMEAN($Rating_{ij}$)	3.4660	19869	.7251	.0051
	$AvgGlobalRating_{ij}$	3.4050	19869	.6020	.0043

Pair 1 SMEAN($Rating_{ij}$) - $AvgGlobalRating_{ij}$

Paired Differences							
			90% Confidence Interval of the Difference				Sig.
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	(2-tailed)
.0611	.5220	.0037	.0550	.0672	16.493	19868	.000

A one-tailed paired samples t-test revealed that, in general for craft beers with 1000 or more ratings, Untappd users gave a more **positive** subsequent rating (i.e., $Rating_{ij}$) compared to the average global rating (i.e., $AvgGlobalRating_{ij}$), $t(19868) = 16.493$, $p < .001$. H1 was accepted.

4.4.2. H2a: Average global rating on subsequent rating in the upper extremity

2a

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SMEAN(Rating$_{ij}$)	4.1557	11324	.4248	.0040
	AvgGlobalRating$_{ij}$	4.1786	11324	.1405	.0013

Pair 1 SMEAN(Rating$_{ij}$) - AvgGlobalRating$_{ij}$

Paired Differences							
			90% Confidence Interval of the Difference				Sig.
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	(2-tailed)
-.0229	.4007	.0038	-.0291	-.0167	-6.077	11323	.000

A one-tailed paired samples t-test revealed that, for craft beers in the upper extremity of the rating scale, Untappd users gave a more **negative** subsequent rating (i.e., Rating$_{ij}$) compared to the average global rating (i.e., AvgGlobalRating$_{ij}$), $t(11323) = -6.077$, $p < .001$. As it was hypothesised that Untappd users would give a more positive subsequent rating compared to the global rating, is was concluded that H2a was not true.

2b

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SMEAN(Rating$_{ij}$)	4.1737	4367	.4276	.0065
	AvgGlobalRating$_{ij}$	4.2043	4367	.1469	.0022

Pair 1 SMEAN(Rating$_{ij}$) - AvgGlobalRating$_{ij}$

Paired Differences							
			90% Confidence Interval of the Difference				Sig.
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	(2-tailed)
-.0306	.4013	.0061	-.0406	-.0206	-5.041	4366	.000

A one-tailed paired samples t-test revealed that, for craft beers in the upper extremity of the rating scale with 1000 or more ratings, Untappd users gave a more **negative** subsequent rating (i.e., Rating$_{ij}$) compared to the average global rating (i.e., AvgGlobalRating$_{ij}$), $t(4366) = -5.041$, $p < .001$. As it was hypothesised that Untappd users would give a more positive subsequent rating compared to the global rating, is was concluded that H2a was not true.

4.4.2. H2b: Average global rating on subsequent rating in the lower extremity

3a

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SMEAN(Rating$_{ij}$)	2.8832	10207	.5962	.0059
	AvgGlobalRating$_{ij}$	2.7136	10207	.2632	.0026

Pair 1 SMEAN(Rating$_{ij}$) - AvgGlobalRating$_{ij}$

Paired Differences							
			90% Confidence Interval of the Difference				
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
.1700	.5595	.0055	.1609	.1791	30.697	10206	.000

A one-tailed paired samples t-test revealed that, for craft beers in the lower extremity of the rating scale, Untappd users give a more **positive** subsequent rating (i.e., Rating$_{ij}$) compared to the average global rating (i.e., AvgGlobalRating$_{ij}$), $t(10206) = 30.697$, $p < .001$. As it was hypothesised that Untappd users would give a more negative subsequent rating compared to the global rating, is was concluded that H2b was not true.

3b

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SMEAN(Rating$_{ij}$)	2.8853	6369	.6064	.0076
	AvgGlobalRating$_{ij}$	2.6997	6369	.2782	.0035

Pair 1 SMEAN(Rating$_{ij}$) - AvgGlobalRating$_{ij}$

Paired Differences							
			90% Confidence Interval of the Difference				
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
.1856	.5646	.0071	.1739	.1972	26.232	6368	.000

A one-tailed paired samples t-test revealed that, for craft beers in the lower extremity of the rating scale and 1000 or more ratings, Untappd users give a more **positive** subsequent rating (i.e., Rating$_{ij}$) compared to the average global rating (i.e., AvgGlobalRating$_{ij}$) for craft beers, $t(6368) = 26.232$, $p < .001$. As it was hypothesised that Untappd users would give a more negative subsequent rating compared to the global rating, is was concluded that H2b was not true.

5. Conclusions, discussion, limitations, implications and future research

In this chapter, conclusions are drawn regarding the research questions, based on the data and literature. Subsequently, it is discussed upon the conclusions. Next, the limitations and implications of this research are discussed. Lastly, recommendations for future research are presented.

5.1. Conclusions

eWOM has a significant impact in online communities, which paves the way for people to be influenced by reviews and ratings of others in these online communities. Both ecommerce and social network research reported a direct effect of prior ratings on subsequent ones through eWOM expressions. Park and Lee (2008) posited two cognitive routes to be influencing consumer decisions, while Brown et al. (2007) focused on consumers' connectedness to others using the online platform and the platform itself, which ultimately leads to a herding effect in these platforms.

This research showed that, in general, subsequent ratings are more positive than the average global rating, indicating a positive herding effect on the Untappd platform. The herding effect is strengthened by the large R^2-value for the average global rating on subsequent rating. More specifically, the average rating in the sample of 50 craft beers in each category is 3.54, which is more than one should expect when a rating scale ranges from 0 to 5. These findings mean that Untappd user's perception about craft beers is positive, which could possibly skew the real quality of the product.

It was found that ratings in rating category "low" deviated more from the mean than ratings in rating category "high" by almost .20 standard deviations, indicating more inconsistency in the rating pattern for lower rated craft beers. Interestingly, the trends in the extremities of rating scales were different than predicted. This research hypothesised that craft beers with an extreme negative (positive) average global rating would also get a more negative (positive) subsequent rating. However, both hypotheses were not true, and the opposite effect was found. This effect could be partly explained by the positive rating tendency on the Untappd platform.

5.2. Discussion

This research found similar results to Muchnik et al. (2013). The positive rating tendency could be attributed to purchasing bias as described by Hu et al (2009), who showed that people that are motivated to buy a product tend to report more positively about it. Furthermore, it could be the case that there are way less experience people on the platform than real experienced ones, which means that the probability to get a more positive rating is greater than getting a more negative subsequent rating (Chorley et al., 2016; Muchnik et al., 2013).

According to Muchnik et al. (2013), a positive rating tendency on a platform would lead to neutralisation, which is found in the subsequent ratings for craft beers with low ratings. Similar results were found by Brown et al. (2007), who stated that the platform itself influences rating behaviour. However, for craft beers with high ratings, it was found that the subsequent rating was more negative than the average global rating. One could argue that this is also the effect of neutralisation, because high ratings would deviate too much from the mean rating average. Adaptation level theory would confirm this, as it posits that consumers evaluate new product experiences according to the standard, the adaptation level (Oliver, 1980). In the case of Untappd ratings, the adaptation level is located in rating category "middle", possibly at 3.54. New user ratings would therefore deviate from this level, which ensures that low ratings get more positive, and high ratings more negative. Moreover, the rater's experience would also explain differences in the extremes. As was posited throughout this research, it could be the case that high rated craft beers are more frequently rated by more experienced people, leading to more negative subsequent ratings (Wang et al., 2018; Ma et al., 2013; Muchnik et al., 2013). When comparing the standard deviations, it was found that low rated craft beers are almost .20 standard deviations further away from high rated craft beers. This could be the case because rating category "low" is much larger than rating category "high". However, it should also be considered that this is the effect of rater's experience, as low rated craft beers could have a more mixed audience of experienced and unexperienced drinkers, because of its availability, price, etc. (Chorley et al., 2016; Muchnik et al., 2013).

When focusing on the actual product experience by users, scholars argued that quality expectancies derived from Untappd ratings could skew consumer's quality experiences (Anderson, 1973). For example, assimilation-contrast theory posits that consumers have perceptual zones of acceptance and rejections. If the difference between expectancy and actual experience is small, they will tend to accept and rate the product more in line with the expected performance. However, when the difference is big, one will reject and magnify the difference and rate the product different from the expected performance. When applying this theory to the findings to this research, it could argued that although low rated craft beers are generally rated low as well (assimilation), they are "not as bad" as expected according to the consumer's adaptation level, and rated more positively to the average global rating (contrast). However, for the high rated craft beers, that are also rated generally high (assimilation), they might be worse than expected according to the consumer's adaptation level and are therefore rated more negatively compared to the average global rating (contrast).

5.3. Limitations and implications

This research assumed a herding effect in Untappd ratings based on prior research of other (beer related) platforms. Besides, one could argue that mean rating that deviates from the real mean of the rating scale and a large R^2-value indicate herding. However, it is acknowledged that there is a lack of hard statistical evidence that confirms the herding effect in Untappd ratings. Claims in this research would be strengthened or weakened in subsequent herding effect research for Untappd specifically.

Based on the literature, it was assumed that positive herding would be present in extreme positive average ratings (H2a), and negative herding in extreme negative ratings (H2b). However, these assumptions turned out to be not true. Given that differences in the extremities of the rating scale are found more often (e.g. Ma et al., 2013; Wang et al., 2018), possible factors that would influence the direction should have been more rigor considering these assumptions. This could have been done by for example first conducting a study about the overall sentiment of people using Untappd.

This research showed some interesting implications for both research and managers. Especially since past researchers have failed to determine the direction of herding effects in online ratings. Given that the hypotheses in the extremities of the rating scales were different than predicted, it is still unclear what these factors actually are. This research, however, draws a theoretical understanding from existing research that factors such as rating tendency of the platform (Muchnik et al., 2013; Brown et al., 2007), user's motivation level (Hu et al., 2009), user's expertise (Wang et al., 2018; Ma et al., 2013; Muchnik et al., 2013), user's adaptation level (Oliver, 1980), and the user's acceptance-rejection zone (Anderson, 1970) to be influencing the direction of ratings on online community platforms. These factors should be considered when conducting rating direction research on online (community) platforms.

For managers, there are some very important implications affecting their business decisions when their products are present in apps like Untappd. First, by default, people rate craft beers quite positive, meaning that they are happy with the craft beers they are drinking. Those positive ratings can thus be used as a marketing tool to show other consumers who are trying to decide which craft beer to buy. Second, having a low average global rating, which could be seen as a bad thing, does not mean that it ends up bad over time. According to the data scraped from Untappd, low rated craft beers get more positive over time. Possibly, those low rated beers can be used as low-priced "bait", to show customers the quality they can get for that particular price. Third, companies that have very high rated craft beers would see that rating descending, as people tend to be disappointed about the quality. Given these implications, it can be recommended to companies with craft beers with low ratings to continue selling that craft beer, as ratings get more positive. Contrary, it is recommended to companies with craft beers with high ratings to work with batches for that particular craft beer (e.g. "Weizen Batch 2") before a negative trend emerges. The latter is especially important, as new batches of a craft beer can be checked-in again by consumers who really liked it. This provides craft beer companies the opportunity to use the same or similar recipe again while still enjoying getting new high ratings for that particular craft beer.

Finally, it is questionable to what extent a herding effect can be controlled for and whether it is needed to control it. Obviously, craft beer companies may set the standard of their ratings by first introducing their craft beer to consumers who they think they will like the beer. This was also shown by Lederrey and West (2018). However, the big question is whether this would be beneficial at all times, as subsequent ratings could be subject to disappointment and negative WOM. It is therefore very important for craft beer companies to make the right assessment and judge what a certain rating means for both the product and its customer.

5.4. Future research

Whilst answering questions of this research, new research questions emerged. For example, only a small fraction of Untappd data is scrapable at the moment. What possibilities would Untappd provide researchers when granting them access to all data they are collecting? This would provide an opportunity to investigate issues in numerous research areas from alcoholism to drinking habits between consumer cultures.

More specifically on consumer behaviour research, what factors ensure the positive rating tendency of a community rating platform? In this research, ratings were divided in three groups “low”, “middle” and “high” (Chorley et al., 2016). Does this really mean that craft beers in category “low” are perceived as bad? And craft beers in category “high” as really great? Are craft beers in category “middle” mediocre? What is considered a good rating? As there was a difference made between “low”, “middle” and “high”, does it also mean that 4 is perceived as very different from 4.1 or even 4.01? What about the difference between 3.9 and 4.0?

Besides users basing their rating on the average global rating, what are the other confounding factors? In this research, average friend rating and even the own average user rating were not taken into account. It would be interesting for future research to see the influencing impact of these factors. It would also be interesting to see the rating tendency of craft beers that are highest or lowest rated or even ratings from established brands.

Finally, do online community platforms such as Untappd have a substantial impact on craft beer sales? This is especially important for craft beer companies that are considering putting a lot of effort in the platform.

6. References

Anderson, R. E. (1973). Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of marketing research*, *10*(1), 38-44.

Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer research*, *14*(3), 350-362.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, *21*(3), 2-20.

Chorley, M. J., Rossi, L., Tyson, G., & Williams, M. J. (2016). Pub crawling at scale: tapping Untappd to explore social drinking. *Tenth International AAAI Conference on Web and Social Media*.

Debnath, S., Mitra, P., & Giles, C. L. (2005). Automatic extraction of informative blocks from webpages. In *Proceedings of the 2005 ACM symposium on Applied computing*, 1722-1726.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science*, *49*(10), 1407-1424.

Duhan, D. F., Johnson, S. D., Wilcox, J. B., & Harrell, G. D. (1997). Influences on consumer use of word-of-mouth recommendation sources. *Journal of the academy of marketing science*, *25*(4), 283.

Esch, F. R., Langner, T., Schmitt, B. H., & Geus, P. (2006). Are brands forever? How brand knowledge and relationships affect current and future purchases. *Journal of product & brand management*.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Gao, G. G., Greenwood, B. N., Agarwal, R., & McCullough, J. (2015). Vocal minority and silent majority: How do online ratings reflect population perceptions of quality?. *Mis Quarterly*, *39*(3), 565-589.

GCA Strategy (2019). Beer still number one on drinks list. Available online: https://www.cga.co.uk/2019/08/01/beer-still-number-one-on-drinks-list-in-great-britain-and-united-states-but-the-challenge-is-engaging-younger-drinkers/

Glenski, M., & Weninger, T. (2017). Predicting user-interactions on reddit. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 609-612. ACM.

Godes, D., & Silva, J. C. (2012). Sequential and temporal dynamics of online opinion. *Marketing Science*, *31*(3), 448-473.

Goswami, A., & Kumar, A. (2019). Online Social Communities. In *Digital Business*, 89-341. Springer, Cham.

Guo, B., & Zhou, S. (2016). Understanding the impact of prior reviews on subsequent reviews: The role of rating volume, variance and reviewer characteristics. *Electronic Commerce Research and Applications*, *20*, 147-158.

Hajli, M. N. (2014). A study of the impact of social media on consumers. *International Journal of Market Research*, *56*(3), 387-404.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?. *Journal of interactive marketing*, *18*(1), 38-52.

Herr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *Journal of consumer research*, *17*(4), 454-462.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, *52*(10), 144-147.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, *60*(11), 2169-2188.

Kwark, Y., Chen, J., & Raghunathan, S. (2014). Online product reviews: Implications for retailers and competing manufacturers. *Information systems research*, *25*(1), 93-110.

Lai, L. S., & Turban, E. (2008). Groups formation and operations in the Web 2.0 environment and social networks. *Group Decision and negotiation*, *17*(5), 387-402.

Lau, G. T., & Ng, S. (2001). Individual and situational factors influencing negative word-of-mouth behaviour. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, *18*(3), 163-178.

Lederrey, G., & West, R. (2018). When sheep shop: measuring herding effects in product ratings with natural experiments. *Proceedings of the 2018 World Wide Web Conference*, 793-802.

Lee, Y. J., Hosanagar, K., & Tan, Y. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science*, *61*(9), 2241-2258.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456-474.

Li, X., & Wu, L. (2018). Herding and social media word-of-mouth: Evidence from Groupon. *Forthcoming at MISQ*.

Ma, X., Khansa, L., Deng, Y., & Kim, S. S. (2013). Impact of prior reviews on the subsequent review process in reputation systems. *Journal of Management Information Systems*, *30*(3), 279-310.

Meuter, M. L., McCabe, D. B., & Curran, J. M. (2013). Electronic word-of-mouth versus interpersonal word-of-mouth: are all forms of word-of-mouth equally influential?. *Services Marketing Quarterly*, *34*(3), 240-256.

Modor Intelligence (2019). Craft Beer Market - Growth, Trends and Forecast (2020 - 2025). Available online: https://www.mordorintelligence.com/industry-reports/craft-beer-market

Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, *31*(3), 372-386.

Moe, W. W., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, *48*(3), 444-456.

Money, R. B., Gilly, M. C., & Graham, J. L. (1998). Explorations of national culture and word-of-mouth referral behavior in the purchase of industrial services in the United States and Japan. *Journal of marketing*, *62*(4), 76-87.

Muchnik, L., Aral, S., and Taylor, S. (2013). Social influence bias: A randomized experiment. *Science, 341*(6146), 647–65.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, *17*(4), 460-469.

O'Reilly, T. (2005). What is web 2.0.

Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, *7*(4), 386-398.

Reingen, P. H., & Kernan, J. B. (1986). Analysis of referral networks in marketing: Methods and illustration. *Journal of marketing research*, *23*(4), 370-378.

Rogers, E. M. (1983). *Diffusion of innovations*. New York: Free Press.

Richins, M. L., & Root-Shaffer, T. (1988). The Role of Involvement and Opinion Leadership in Consumer Word-of-Mouth: An Implicit Model Made Explicit. *Advances in Consumer Research*, *15*(1).

Schachter, S. (1959). The psychology of affiliation: Experimental studies of the sources of gregariousness.

Schlosser, A. E. (2005). Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*, *32*(2), 260-265.

Shankar, V., Venkatesh, A., Hofacker, C., & Naik, P. (2010). Mobile marketing in the retailing environment: current insights and future research avenues. *Journal of interactive marketing*, *24*(2), 111-120.

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

Statista (2020). Mobile operating systems' market share worldwide from January 2012 to December 2019. Available online: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/

Sun, T., Youn, S., Wu, G., & Kuntaraporn, M. (2006). Online word-of-mouth (or mouse): An exploration of its antecedents and consequences. *Journal of Computer-Mediated Communication*, *11*(4), 1104-1127.

Thakur, R. (2016). Understanding customer engagement and loyalty: a case of mobile devices for shopping. *Journal of Retailing and consumer Services*, *32*, 151-163.

Tucker, C., & Zhang, J. (2011). How does popularity information affect choices? A field experiment. *Management Science*, *57*(5), 828-842.

Untappd (n.d.). Businesses. Available online: https://untappd.com/business

Untappd (2019). Year In Beer 2019. Available online: https://blog.untappd.com/post/189742210691/year-in-beer-2019

Venkatesan, M. (1966). Experimental study of consumer behavior conformity and independence. *Journal of marketing research*, *3*(4), 384-387.

Wang, C., Zhang, X., & Hann, I. H. (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research*, *29*(3), 641-655.

Wilsnack, R. W., Wilsnack, S. C., Kristjanson, A. F., Vogeltanz-Holm, N. D., & Gmel, G. (2009). Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addiction*, *104*(9), 1487-1500.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS quarterly*, *38*(2), 539-560.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*(2), 133-148.

7. Appendix

7.1. List of randomised numbers

The numbers are organised from top to bottom, then the next column. Subsequently, they will continue on the next page. The same reading method applies.

821	111	898	335	536	458	332	121	421	808	480	744	281
150	53	38	208	462	54	119	459	441	260	592	800	217
593	542	660	617	642	599	773	343	330	710	305	234	138
602	794	319	13	842	633	366	388	87	806	555	454	576
641	455	219	381	733	645	846	652	549	155	508	130	7
345	848	299	456	282	339	644	215	581	187	375	160	389
586	302	532	519	863	790	805	827	6	638	352	860	440
889	715	493	257	285	292	618	104	554	600	812	828	580
492	899	12	84	107	200	382	878	337	436	356	699	183
128	48	443	807	71	93	877	768	274	587	378	14	403
425	193	338	224	533	717	721	619	369	173	678	171	126
61	611	267	120	484	575	596	833	627	146	165	152	730
713	825	278	547	422	475	551	795	603	91	497	639	279
590	496	15	803	743	683	502	896	137	358	2	326	722
814	767	538	16	757	19	753	231	207	158	529	216	166
248	390	359	65	280	791	637	445	50	885	289	370	653
453	588	22	731	650	482	324	857	10	434	838	309	861
582	630	396	333	272	379	153	360	572	325	244	201	32
363	649	427	98	76	854	449	525	213	672	43	398	598
688	420	747	321	524	507	875	129	671	859	428	313	616
136	594	145	349	103	196	451	255	840	99	56	20	585
545	716	59	387	141	70	696	78	635	189	386	819	431
197	293	725	726	781	433	361	697	241	465	66	690	86
798	742	320	407	523	34	736	874	568	584	276	392	51
488	27	615	631	72	77	620	397	307	573	85	740	385
55	290	410	376	628	416	685	494	264	112	612	674	181
756	546	578	294	832	194	661	336	461	391	900	423	552
314	495	509	457	739	728	221	400	206	897	254	510	227
177	565	686	853	100	101	30	28	718	782	858	41	471
541	720	837	371	180	895	142	253	830	748	481	182	687
368	535	223	591	295	809	707	401	681	214	23	24	265
662	64	113	732	701	579	143	867	831	505	315	864	792
301	82	780	198	62	97	629	526	569	261	574	665	133
220	308	342	749	115	882	235	777	437	222	88	677	123
277	556	539	799	478	92	559	865	751	774	105	700	340
811	499	849	409	621	394	75	689	35	268	503	300	122
419	779	159	845	147	479	564	334	787	418	659	435	511
657	474	186	706	446	116	823	447	191	250	402	513	784
553	643	31	452	463	21	640	668	892	4	364	469	467
408	550	761	17	527	658	655	49	543	383	664	734	834
729	317	483	262	413	466	804	868	663	520	605	597	500
824	624	506	323	836	537	760	448	785	850	384	560	8
44	83	862	271	374	517	632	225	249	304	684	741	184

169	210	328	852	566	610	162	521	829	11	341	47	680
242	218	886	39	856	702	167	765	797	719	57	745	883
567	205	766	170	589	232	259	40	230	601	772	881	74
822	273	693	296	646	666	548	583	178	355	450	816	470
411	96	269	634	623	161	873	698	90	240	179	692	188
818	786	491	673	412	604	33	647	727	490	134	426	464
595	344	738	175	399	709	891	58	669	531	820	168	26
251	439	67	626	156	703	810	211	704	373	195	270	163
872	393	288	199	291	176	622	79	238	275	851	855	712
185	312	750	164	512	42	63	844	245	522	258	514	430
172	438	174	752	485	705	233	94	102	802	247	3	118
763	380	60	212	711	347	775	89	884	298	835	762	724
80	372	534	460	229	414	18	759	424	69	871	486	656
876	577	625	45	365	110	764	331	670	149	125	888	769
562	311	737	417	286	226	429	68	151	771	135	310	109
353	246	636	691	788	894	350	395	132	606	557	148	675
204	609	890	648	893	758	778	318	571	477	52	530	776
9	826	561	528	708	714	841	203	815	793	362	131	
377	654	607	263	367	243	190	839	106	473	1	676	
746	228	570	237	879	127	192	306	801	518	37	843	
476	813	869	73	735	444	144	351	284	682	432	287	
348	95	154	651	252	516	5	558	406	870	283	754	
694	442	472	540	157	108	81	297	139	303	667	613	
114	415	679	498	887	789	468	504	489	209	608	117	
29	755	236	239	723	46	796	783	124	316	346	817	
501	354	770	544	202	563	329	880	266	695	357	322	
614	327	866	256	515	36	404	25	847	140	405	487	

7.2. Missing values in dataset

Missing Data

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No values missing	34824	93.9	93.9	93.9
	One value missing	2253	6.1	6.1	100.0
	All values missing	3	.0	.0	100.0
	Total	37080	100.0	100.0	

7.3. Check for relationship

Group Statistics

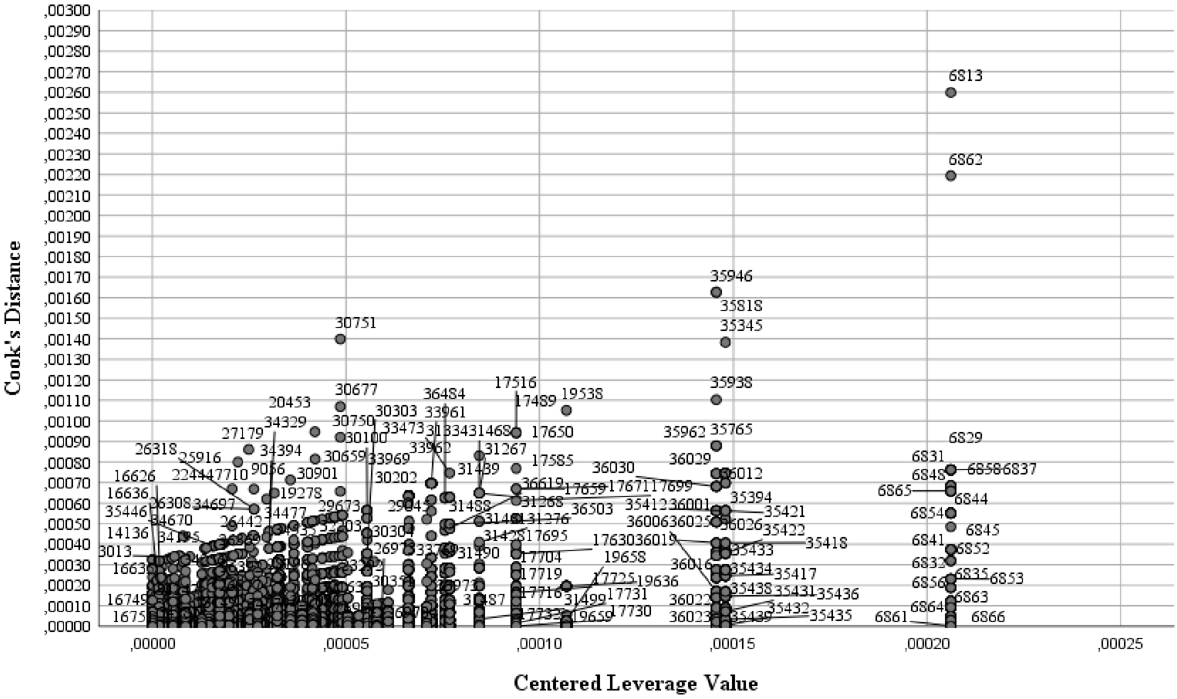
	Missing User Rating	N	Mean	Std. Deviation	Std. Error Mean
AvgGlobalRating$_{ij}$	.00	2253	3.4145125	.63825421	.01344662
	1.00	34824	3.4370277	.64767217	.00347069
Rating$_{ij}$	.00	0[a]	.	.	.
	1.00	34824	3.4704	.88829	.00476

a. t cannot be computed because at least one of the groups is empty.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Rating$_{ij}$	Equal variances assumed	2.516	.113	-1.601	37075	.109	-.02251518	.01406716	-.05008720	.00505684
	Equal variances not assumed			-1.621	2561.317	.105	-.02251518	.01388731	-.04974667	.00471630

7.4. Scatter of Cook's Distance by Centered Leverage Value



7.5. Regression output after treatment

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.734[a]	.539	.539	.49768	.539	40478.606	1	34672	.000

a. Predictors: (Constant), $AvgGlobalRating_{ij}$

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10025.853	1	10025.853	40478.606	.000[b]
	Residual	8587.657	34672	.248		
	Total	18613.510	34673			

a. Dependent Variable: SMEAN($Rating_{ij}$)

b. Predictors: (Constant), $AvgGlobalRating_{ij}$

Coefficients[a]

Model		Unstandardized Coefficients: B	Std. Error	Standardized Coefficients: Beta	t	Sig.	95,0% Confidence Interval for B: Lower Bound	Upper Bound
1	(Constant)	.515	.015		33.670	.000	.485	.545
	$AvgGlobalRating_{ij}$	.868	.004	.734	201.193	.000	.860	.877

a. Dependent Variable: SMEAN($Rating_{ij}$)