Influencing perceived consumer experiences, before and after they have happened

A quantitative study of ratings influence on perceived consumer experience
in both a pre- and post-experience setting

Eleonor Toni (23230) & Erik Nilestoff (50316)

Abstract: Academia has for a long time studied the influence of word of mouth, specifically reviews, on consumers' purchase intentions of products and the transformative effects thereof. Managers use this knowledge heavily to their advantage, i.e. to attract new customers. Ratings, i.e. the aggregated review, are increasingly being employed e.g online and in mobile apps, and we are expected to answer them more often. It is hard to escape them as we are faced with ratings wherever we go and whatever we do, both provided by the company itself and third party platforms. We know that ratings that we receive increase our purchase intentions - but how do they relate to the actual perceived consumer experience, i.e. how consumers rate their own experience, their perceived satisfaction and fulfillment of expectations on the experience? Using two experimental studies, i.e. a taste experience and a film experience, we managed to show that ratings can influence consumers' perceived experiences both if the rating thereof is received before living the experience and after, i.e. when the consumer has already had their chance to form an opinion about it. The exposure to a rating moves the evaluation of the experience in the direction of the rating, i.e. a low rating lowers the evaluation and vice versa. The findings provided are of high value and importance for both theoretical and managerial use.

Keywords: influence of reviews, forward-framing, backward-framing, consumer satisfaction, consumer experience

Acknowledgements

Our sincerest thanks to

our supervisor **Micael Dahlen**
for knowledge, support and ceaseless enthusiasm throughout the process

Head of Analysis **Robin Fischer** at Filmstaden
for bouncing ideas with us and making a reality of the film experiment

our fellow student **Zacharias Lindqvist Hansson**
for referring us to Filmstaden

Marketing Director **Anna Bartholf** at Cloetta
for sponsoring us with chocolate needed for the taste experiment

all chocolate respondents in the streets of Stockholm
that took the time to taste the chocolate and answer our questionnaire

all filmgoing respondents in the Filmstaden panel
that after seeing the film took the time to answer our questionnaire

family and friends
that has supported us with insights, referrals and been cheering us on

the faculty of Stockholm school of Economics
that provided us with the tools and knowledge to write this thesis

Table of definitions

Pre-experience vs post-experience setting (Adapted to the context of this thesis from Braun, 1999)	The temporal setting when stimulus information is given to respondents, i.e. *before* or *after* an experience, respectively. Connected to the concepts of forward-framing and backward-framing.
Forward-framing (Adopted to the context of this thesis from Braun-LaTour & LaTour, 2005)	Stimulus information about an experience presented to you *before* going through the experience, framing you to evaluate it accordingly.
Backward-framing (Adopted to the context of this thesis from Braun-LaTour & LaTour, 2005)	Stimulus information about an experience presented to you *after* going through the experience, framing you to reconstruct your own memories about the experience and evaluate it accordingly.
Anchoring effect (Adopted to the context of this thesis from Furnham & Boo, 2010)	Stimulus information received *before* going through an experience, that provides mental cues, i.e. "anchors" in the consumers' mind, motivating them to evaluate their perceived experiences in the direction of the anchored information.
Primacy effect (First described by Fredrick Hansen Lund, 1925)	Persuasion arguments presented *first* rather than *subsequently*, will have greater effectiveness in persuasion.
Recency effect (First described by Cromwell, 1950)	Persuasion arguments presented *later* rather than *first*, will have greater effectiveness in persuasion.
Perceived consumer experience See **section 2.2** for the theoretical basis of this definition.	**Our derived definition:** Consumers' evaluation of their experience on the four following variables: ● Fulfillment of expectations ● Satisfaction ● Rating (of their own experience) ● Purchase intentions.
Variable: Fulfillment of expectations (Oliver, 1980)	Fulfillment of expectations is a variable of "the degree to which the product (also service or experience) exceeds, meets, or falls short of one's expectations"
Variable: Rating (Qiu, Pang & Lim, 2012)	Reviews are divided into two types, individual reviews and aggregated rating. With "rating" we mean the aggregated rating. In previous research, the word "evaluation" is sometimes used interchangeably with "rating", whereby we use "evaluation" when referring to such literature but we use "rating" when referring to our variable.
Variable: Satisfaction (Jones, Mothersbaugh & Betty, 2000)	Satisfaction is a variable that measures an overall evaluation of performance for products or services.
Variable: Purchase intentions (Chapman & Aylesworth, 1999)	Purchase intentions is a variable for measuring the extent to which a consumer intends to buy a product or service.

Table of contents

1. Introduction

Reviews and ratings - "if it got five stars out of five I can be sure that I am going to have a good time, right?". This section aims to outline the overall process of how we derived our research question. While the background section is more aimed towards demonstrating the timeliness of this research and waking the interest of the reader, as it did us, the subsequent section will present the purpose of this thesis and argue for our expected contribution to research.

1.1. Background

Living in a world in which consumers are increasingly reliant upon reviews and ratings in decision-making processes, and companies using them to maintain their brand image and up their sales, we are faced with reviews and ratings basically everywhere we go. In restaurants, shops and after going to the theatre, yes, even after taking a taxi ride home from the concert you just went to, you are faced with demands of rating or reviewing your experience with the driver in question. When ordering your taxi you see that the driver who is picking you up has a rating of 4.9 out of five, could it be that this ride is going to be outstanding? Well, all taxi drivers seem to have at least 4.5 stars and since they tell you they will give you, as a client, five stars, you give them the same, with little reflection over the possible effects of your decision to do so. Would you have done that if you had not yet received the information that they will give you five stars out of five? This particular phenomenon could most likely be explained by the principle of reciprocity (Cialdini, 1987). Also, recent scientific literature highlights that rating scores are becoming inflated (O'Connor & Cheema, 2018) the more they are used. After giving the taxi driver a 5-star rating for providing an average ride home you think back to that concert you just saw and how you thought it was an unforgettable 5-star experience. Before going to bed, you browse your favorite online newspaper only to discover that the majority thought the concert was overrated and it only received a 3.5 rating. The next day at work your colleague ask you how the concert was, you think back upon the experience and answer your colleague: "It was alright".

So what is the purpose of this thesis? The taxi and concert examples is what made us look into previous literature concerning the effects of reviews and ratings. We wondered how reviews and ratings make us reflect upon an upcoming or past experience and how other people have rated their experience influence yours. In the following section we will give an introductory overview of what we found and state what research question this resulted in, as well as how we aim to contribute to research in the field of ratings and reviews.

1.2. Purpose, expected contribution & research question

A lot of literature has been written based on research done concerning the effects of reviews in particular, as we will show in the following chapter. However, much of this research is focused towards the effects on attitudes towards the brand or on purchase intentions and how this relates to sales (e.g. Hilger, Rafert & Villas-Boas, 2011; Vermeulen & Seegers, 2009; Yang, 2016; Zhu & Zhang, 2006). These findings are of course of importance for literature, but maybe their greatest relevance are for managers seeking to up their companies' sales volumes, through the use of reviews as a kind of advertising information (given that their companies and their products are positively evaluated in reviews or given high ratings). For the purpose of this thesis, we will argue for that an often disregarded measure in the field concerning the influence of reviews is that of *satisfaction*, which is associated with purchase intentions (Oliver, 1980), and to which we will argue that customers' own *rating* of their experiences as well as the perception of the extent to which their *expectations are fulfilled* are highly linked. This thesis is therefore of high phenomenological importance. We will use these three measures that we will define as key indicators of change in *perceived consumer experience*s together with literature concerning the *temporal setting* in which reviews have been shown to have an influence. By using manipulated (aggregate) *ratings*, which is a type of review (Qiu, Pang & Lim, 2012), the goal of this thesis is to answer the research question that follows below. Given that managers not only intend to increase purchase intentions, but also increase the satisfaction of their customers, the thesis are also of high managerial importance.

Research question: *Can perceived consumer experiences be influenced by ratings both in a pre- and post-experience setting?*

1.3. Delimitations

While reviews are given and employed as a source of information for various product categories in a global setting, this thesis is delimited to Sweden and the first study to Stockholm where the sample was collected. The second study is delimited to panelists among Filmstadens filmgoers. These delimitations were due to limited resources and access to respondents. Concerning the offered product experiences, these were delimited to a chocolate bar tasting experience and a film experience. Although these two represent two rather different product categories, which may be argued to increase the generalizability of the results across product categories, this study is still delimited to those same product categories. Furthermore, the chocolate tasting study does not take into account the setting in which chocolate is normally consumed, although the chocolate was experienced where the respondents were at the time being, and thus not in a laboratory setting. The film experience study does not take into account that the filmgoers have visited different theatres all over Sweden, with different equipment or with other factors of convenience and comfort in the theatres. Neither does it take into account the different times when the film has been experienced by respondents, nor the amount of time passed between the film experience and the answering of the questionnaire. As the respondents in the chocolate experiment needed to answer the questionnaire precisely before and after they had tasted the chocolate, the potential effect of having had more time to reflect about the experience is beyond the scope of this thesis. Furthermore, this thesis is delimited to reviews in the aggregate form, i.e. number-based ratings, and their potential effect on the variables that we will argue for constituting the "perceived consumer experience" in the following chapter.

2. Theoretical framework

We use existing theory to outline the derivation of our research question and hypotheses. By splitting the research question into three parts, we explain how we built the theoretical foundation upon which the research question stands.

2.1. Theoretical approach

The research question: "Can perceived consumer experiences be influenced by ratings in both a pre- and a post-experience setting?" suggests that ratings may have an influence that leads to a change within the consumer's minds about their own perceived experience, i.e. how consumers rate their own experience, their perceived satisfaction and fulfillment of expectations on it as well as their purchase intentions. In order to describe how we derived this question, we split it into three parts. Firstly, we needed to define the change per se, i.e. defining what we mean with a change in "the perceived consumer experience". Secondly, we needed to define w*hat* induces the change within the consumer's minds, i.e. defining "the influence of ratings". Thirdly, we needed to define *when* we expected the change of the perceived consumer experience to take place, i.e. explaining the inclusion of "in both a pre- and post-experience setting" into the research question. Defining *why* this change may occur at these points in time is not demanded by the formulation of the research question, but we partly included this in order to further strengthen the justification of our hypotheses.

2.2. Defining perceived consumer experiences

The change we have hypothesized is in the (consumer's) perception of the consumer experience. To be able to determine whether or not this change would occur, we needed to define what we mean with "the perceived consumer experience" and what variables this definition includes, in order to be able to employ them as key indicators of change in perceived consumer experiences in our tests.

2.2.1. Purchase intentions

As the influence of reviews often is described to take place in purchase intentions (e.g. Hilger, Rafert & Villas-Boas, 2011; Vermeulen & Seegers, 2009; Yang, 2016; Zhu & Zhang, 2006), and similarly for the influence of word-of-mouth, WOM (Trusov, Bucklin & Pauwels, 2009), i.e. the passing of information from one person to another through oral communication, we will use the variable *purchase intentions* as a starting point to define what we mean with "the perceived consumer experience" and therefore also include it as a key indicator of change.

2.2.2. Satisfaction

Satisfaction is a recurrent measure used in academic literature when it comes to evaluating products, service or experiences (Jones, Mothersbaugh & Betty, 2000). It is often seen as an overall post-purchase evaluation of products or services, affected by and correlated with other pre-purchase and during-purchase variables, such as attitude change and purchase intentions (Oliver, 1980). Arguments whether purchase intentions can be measured has been made by researchers, however it has been shown that satisfaction affect purchase intentions positively (Cronin & Taylor, 1992). Satisfaction is therefore an important part of the consumer experience and a relevant measure for this study.

2.2.3. Rating

The availability, quantity and access to reviews has in the last couple of years increased replacing other previously used information such as product description (Mudambi & Schuff, 2010). Reviews are divided into two types, individual reviews and aggregated rating (Qiu, Pang & Lim, 2012). This thesis focuses on the effects aggregated ratings has on consumers perceived experience and rating is therefore used as both an influencing variable and a relevant measure of outcome.

2.2.4. Fulfillment of expectations

Consumer satisfaction is in turn a function of expectations and expectancy and the expectations consumers have on products, services and experiences can either be confirmed

or disconfirmed according to the expectancy-disconfirmation framework (Oliver, 1980). How well the expectations of the consumers are fulfilled is therefore a crucial part of the perceived consumer experience and is an obvious key measure for this study.

2.3. The influence of WOM, reviews and ratings

As we defined ratings as aggregated reviews and reviews in turn can be argued to be a form of online word-of-mouth, i.e. eWOM, the following section will serve to delineate what research have previously been conducted concerning the influence of word-of-mouth, reviews and ratings. The aim is to derive the hypothesized "influence of ratings", demanded by our research question.

Referrals by word-of-mouth have been shown to have considerably longer carry-over effects than more traditional marketing actions, such as advertising campaigns, and they also produce much higher response elasticities. (Trusov, Bucklin & Pauwels, 2009). Also the effect of reviews are often linked to purchase intentions, consumer choice and/or increased sales volumes (Hilger, Rafert & Villas-Boas, 2011; Vermeulen & Seegers, 2009; Yang, 2016; Zhu & Zhang, 2006). For example, a one-point increase in the average rating of a video game, translates to a 4 % increase in game sales (Zhu & Zhang, 2006). Purchase intentions after receiving a review have also been suggested to differ based on gender, level of subjectivity (or objectivity) and hedonic versus utilitarian context (Liu, Ozanne & Mattila, 2018). Men's purchase intentions in a hedonic context were increased when they were provided with reviews containing subjective expressions, while this impact were larger for women in the utilitarian context (Liu, Ozanne & Mattila, 2018). Exposure to online reviews enhances consideration in consumers (Vermeulen & Seegers, 2009) and the authors suggest that both positive and negative reviews increase consumer awareness, while positive reviews also improve attitudes towards the brand, an effect that was even stronger for lesser-known brands.

Reviews have also been shown to have an effect on the evaluation of an experience, making the consumers evaluate the experience more in line with the review than they would have without being primed with it (Wyatt and Badger, 1984). The extent to which reviews impact

evaluation of a product experience seems to be dependent on the type of product. Consumer evaluation of products with hedonic properties is suggested to be much less influenced by reviews than when products have more utilitarian properties (Cervellon and Carey, 2014). Further dividing the evaluation of utilitarian products into assessments of ambiguous and unambiguous properties of the products, consumers seemed to rely much more on reviews when evaluating ambiguous properties, i.e. properties that they found hard to assess themselves. Reviews thus seem to have a larger impact on evaluation for utilitarian ambiguous products than for utilitarian unambiguous or hedonic products.

When categorizing products into search goods (similar to utilitarian products) and experience goods (similar to hedonic products), the evaluation of the search goods was also suggested to be influenced to a greater extent of positive reviews than the evaluation of the experience goods (Hao, Ye, Li & Cheng, 2010). In the example of the video game mentioned above, Zhu & Zhang (2006) classified the video game as an experience good, whereby we thus assume that a similar increase in the average rating of a search good would translate to an even higher impact on sales volumes. The impact of negative reviews showed no significant difference between these types of goods, however, the difference in impact between negative and positive reviews was greater for experience goods than for search goods (Hao, Ye, Li & Cheng, 2010).

Further supporting that positive reviews seem to have a great impact on evaluation, we also see that reviews conducted in a very enthusiastic manner, so called rave reviews, have been suggested to positively influence the evaluation of the product in question, as well as transferring the positive evaluation onto the parent company and thus other related products within the brand. This attitude transfer also seem to have an effect on related products outside of the brand family, although smaller. (Chapman and Aylesworth, 1999).

For negative reviews, the impact is dependent on the level of involvement of the consumer and the perceived quality of the review. An increasing proportion of negative reviews will impact low-involvement consumers to conform to the opinions representative of the review regardless of the quality of the reviews, while such an increase will only influence high-involvement consumers if they perceive the quality of the reviews to be high. (Lee,

2007). The level-of-involvement depended upon the product type, i.e. customers less involved with film, i.e. interested in and/or knowledgeable about film, would be influenced by negative reviews to a larger extent than those who were more involved. This goes in line with the findings of Hoch & Deighton (1989) that attempted to identify where *learning* from product consumption experience is "most open to managerial influence", and came to the conclusion that consumers' openness to this process is dependent upon three moderating factors that influences learning: the familiarity with the domain, the motivation to learn and the ambiguity of the information environment.

Further supporting that product type moderates the effect of reviews, here: on the perceived helpfulness of the review, are Mudambi & Schuff (2010). They showed that review extremity, review depth and product type affected the perceived helpfulness of the review. Product type (search good or experience good) moderated the effect that review extremity had on perceived helpfulness of a review, where extreme ratings were less helpful than reviews with moderate rating for experience goods. For both products, review depth had a positive influence on the perceived helpfulness of the review, but also here did the product type moderate the effect of review depth on the helpfulness of the review, where the positive effect of review depth was greater on the perceived helpfulness of the review for search goods. These findings also indicate that other factors than product information affects the perceived helpfulness of reviews, which has been further explored in the context of online reviews by authors such as Forman, Ghose & Wiesenfeld (2008) and Schindler & Bickart (2012).

The influence of reviews in online communities has been suggested to be greater if the review includes identity-descriptive information about the reviewer (Forman, Ghose & Wiesenfeld, 2008; Schindler & Bickart (2012), as this information will be used as a complement or replacement of product information augmenting the purchase intentions of the product and the evaluation of the helpfulness of the review. This effect was even larger if the identity-descriptive information revealed that the reviewer was from the same geographical place as the potential customer. (Forman, Ghose & Wiesenfeld, 2008). Review elements impairing clarity, such as spelling mistakes and grammatical errors, was further shown to reduce the perceived helpfulness of the review, while entertaining review elements, such as

expressive slang and humour, increased the same (Schindler & Bickart, 2012). Also, the review length had importance, as moderate length reviews were evaluated as most helpful.

Who wrote the review is also suggested to be important as Hilger, Rafert & Villas-Boas (2011) showed that the demand for a product (in their case: wine bottles) changed in the direction of which experts had rated the wine, indicating that quality information is transmitted through expert opinion labels, in which the score of the rating play a crucial role. Who the receiver of the review is has also been shown to have importance as Wu, Shen, Li & Deng (2017) demonstrates how sharing a temporal contiguity cue when reviewing e.g. a trip, such as "just got back from the trip" enhanced perceived trustworthiness of the review for consumers having a low personal sense of power, while reducing it for consumers having a high sense of personal power. This finding is also in line with Zhu & Zhang (2010) who found that online reviews are less influential on product sales if the consumer has greater internet experience, further emphasizing the role of consumer characteristics.

We thus see that a lot of previous research focus on the influence of reviews, ratings and WOM on purchase intentions, and the different factors of what makes a good review, i.e. one that leads to increased purchase intentions and/or product sales. We have previously argued for that purchase intentions are highly linked to consumer satisfaction and thus also their rating of the experience, as well as their perceived fulfillment of expectations on the experience. However, we believe that is important to state that one might also argue that experiences are highly ambivalent and open for interpretation. For example, Hoch (2002) argues that product experiences are seductive in the way that they make consumers believe that they learn more through the experience than they actually do, and that one reason for this is that an experience is more engaging than other types of education and thus more memorable. For the purpose of this thesis, we will however delimit the scope to the four variables argued for in the previous section.

2.4. The timing of the influence

Ratings can be used as advertisement information. Perceptions about a product or service have been shown to be possible to alter by giving the consumers advertisement information, both in a pre- and post-experience setting. This section therefore serves to outline what has been shown in literature before with respect to the timing of the information given.

In the world of advertising, it is often hard to measure the impact your advertising efforts have on consumers, which has made authors such as Hall (2002) investigate and develop new models for measuring advertising effectiveness. Emotions have been shown to act as a mediator between advertising content and consumer responses, i.e. attitudes towards the ad or the brand (Holbrook & Batra, 1997). Hall (2002) describes how marketers and senior managers all too often apply a thinking process derived from the AIDA model (Attention-Interest-Desire-Action), where the process is linear in the sense that it starts with consumers "changing their minds" about a product, i.e. a cognitive process, after which they change their attitude, meaning that the *cognition* translates into *affect* and finally the consumers act, entailing a translation into *behaviour*.

Hall (2002) means that the problem which such a model is that its primary goal with advertising is to induce trial by the consumer in order to insert the brand in question into their minds and to maintain it there. Vakratas & Ambler (1999) classify advertising effects into two sub-categories: intermediate effects, i.e. on consumer beliefs and attitudes, and behavioral effects, i.e. such that relates to purchase behaviour, e.g. brand choices, but similarly to Hall (2002), the authors also state that there is little support for a hierarchical, temporal sequence in which these effects occur. Hall (2002) however place critique on the extent to which cognition is given a role in consumers' responses to advertising, including the work of both Vakratas & Ambler (1999) as well as Holbrook & Batra (1997) and develops a new model for consumers' responses to advertising. The role of cognition has been reduced greatly and the concept of *perception*, i.e. a dependent variable influenced by advertising and experience, that is included to the model as one out of three key elements. The other two are *experience* and *memory*, and the model comprises a dimension of multiple feedback loops,

connecting advertising and perception at every stage of the process. There are three phases of the model.

In the first phase described by Hall (2002), *the pre-experience exposure* (to advertising) phase, the advertisement frames perception, leading to *expectations* and *anticipation* of the experience, as well as a *rationale* for the anticipation. The rationale is often explicit in terms of product features or other things that are either described or implied as objective reasons to buy it. In the second phase, the advertising information enhances the sensory experience, which has experimentally been shown to occur both in the pre- and post-experience phase, i.e. both when the consumer has been exposed to an advertisement *before* the experience and *after* (Hall, 2002). As Hoch (2002) argued that experiences, by being engaging, seduce consumers into thinking they learn more from the experience than they actually do, we argue that in combination with Halls (2002) findings, this suggests that advertising information enhances the sensory experience, i.e. some engaging elements of the experience and thereby also the seductive element of the experience. In the third phase, *the post-experience exposure* phase, memory is organized, i.e. providing mental cues for the recall of the experience, as well as interpreted, a process in which the advertisement is not only making the consumer feel good about the experience, but also providing explicit or implicit reasons for the consumer to actually believe that it was. Hall (2002) argues that pure pre-experience exposure may only occur prior to launch of a completely new product or among those have not yet tried a product, otherwise the consumer will always be trapped in a cycle where they are "exposed to the advertising in a continuous loop between post- and pre-experience", in which the "distinction between pre- and post- blurs", as the advertising operate on two dimensions, to organize memory of the last experience, while also framing the perception of the next one.

Prior to this research, it had been concluded that advertising may have an effect in both the pre-experience setting (e.g. Hoch & Ha, 1986) and the post-experience phase (e.g. Braun, 1999), but Braun-LaTour, LaTour, Pickrell & Loftus (2004), Braun-LaTour & LaTour (2005) and Braun-LaTour, Grinley & Loftus (2006) also attempted to isolate the effect of advertisement information in the post-experience setting in order to study the phenomenon of "false memory creation" (all three studies) and determine which temporal setting had the

largest impact on the same (Braun-LaTour & LaTour, 2005). For the purpose of this thesis, we will therefore continue to treat these two temporal settings separately and will in the following two sections elaborate further on what has been shown in previous literature, with respect to the two temporal settings.

2.4.1. Forward-framing (information received pre-experience)

Advertisements can be considered "tentative hypotheses" which consumers test through product experience (Hoch & Ha, 1986). When provided with unambiguous (objective and physical) evidence about the product quality, the impact of advertising were insignificant on consumers' judgements of product quality. In the case of ambiguous evidence, advertising influenced quality judgements through a process in which the encoding of the physical evidence were affected and retrieval of evidence consistent with the advertisement seemed to occur (Hoch & Ha, 1986). These findings thus supports pre-experience advertisements' transformative effects on consumer evaluation, when they are provided with ambiguous evidence, but not unambiguous.

Levin & Gaeth (1988) have shown that framing advertising information in a more positive way (i.e. labelling meat as 75% lean instead of 25% fat) led to the consumers rating the product higher on qualitative attributes *before* tasting it. However, this effect was diminished when the consumers actually tasted the meat, suggesting that the forward-framing effect is diluted if the consumer is then faced with a diagnostic product experience.

One might argue that forward-framing may work thanks to *the law of primacy* in persuasion (first described by Fredrick Hansen Lund, 1925). It states that the order in which information is presented influences the formation of opinion around it, i.e. what is presented first, e.g. advertisement information or a review, will be more effective in persuasion than what is presented subsequently, e.g. information gained from "living the experience". Arguments about *the anchoring effect* (first introduced by Slovic, 1967; the notion later reviewed in literature by Furnham & Boo, 2010, for its broad number of perspectives) support this theory, as advertisement information or a review received pre-experience, will "anchor" into the consumer's minds, motivating them to evaluate their perceived experiences in the direction of the anchored information, i.e. the advertisement information or the review. For the purpose of

this thesis, we have thus chosen to employ the perspective that anchors are used as a cue or a hint that influences "the information processing that bias judgements towards the anchor" (Furnham & Boo, 2010). In the middle of the 20th century, the theory about the law of primacy was opposed by that of *the recency effect*, which stated that persuasion arguments presented later rather than first, had greater effectiveness in persuasion (first described by Cromwell, 1950), however nowadays it has been concluded that both of these effects may occur (Castel, 2008; Steiner & Rain, 1989). This leads us into the field of backward-framing, as it may be argued that an argument, e.g. advertisement information or a review, may as effective in persuasion if presented shortly *after* the lived experience, as stated by the recency effect, as *before* it, through the mechanism underlying the law of primacy.

2.4.2. Backward-framing (information received post-experience)

In advertising, post-experience information have been shown to influence the cognitive processes of consumers' minds, thus altering their experiences to the extent that one can start believing that the advertising information is representative of one's own experience. Over time, this has been shown to be incorporated into brand schemas and thus potential to alter future product decisions. (Braun, 1999). The process through which this phenomenon occurs is called a "reconstructive memory process", further explored by the same author in the context of tourist's travelling stories (Braun-LaTour, Grinley & Loftus, 2006). Both advertising information and information given through word-of-mouth were shown to influence or distort tourist's explicit memories. False or misleading information (used a treatment) were in this context shown to not only reconstruct travellers' own personal memories, but to also alter their semantic memory for the destination visited. Treatment repetition of false information increased its memory distorting effect, i.e. lead to a greater false memory creation. False memory creation have also been shown to be greatest when the false information is given pictorially, rather than verbally (Braun-LaTour, LaTour, Pickrell & Loftus, 2004).

Hoch & Ha (1986) had, as previously explained, shown that the transformational effect of advertising is dependent upon the retrieval of evidence consistent with the advertisement, and the authors also saw a greater increase in the average rating given by the group that received the advertisement information *beforehand* than in the average rating given by those who saw

it *after* their experience, suggesting that this effect was due to the after-group not being able to retrieve as much information from *the advertisement* consistent with that of *their lived experience*, as the before-group retrieved from *their lived experience* consistent with *the advertisement*. Braun-LaTour & LaTour (2005) later showed that advertising information given shortly after the consumer experience had a larger transformational impact on the memory of it, due to facilitating consumers' memory recall, than if given beforehand, as it in the latter case would rather "contaminate the memory". As retrieval of information from an advertisement was selective in the sense that consumers tend to retrieve only the information that they found consistent with their experience, or vice versa (Hoch & Ha, 1986), we argue that consistency theories are relevant for the purpose of this thesis. Consistency theories are built upon the assumption that people are motivated to seek coherence and counteract cognitive dissonance in order to create consistency within their minds, i.e. among their thoughts (Cialdini, 1987). This implies that people with inconsistent thoughts will relieve inconsistency by altering their thoughts in the direction of which they perceive their thoughts to be more consistent. For the purpose of this study, this further implies that consumers' perceived experiences, would change in the direction of the rating received, both in a pre- and post-experience setting, as inconsistency otherwise would be present.

2.5. Research question

As we have argued with basis in literature that there are dependent variables that reviews are likely to affect, i.e. satisfaction, rating and fulfillment of expectations, other than purchase intentions and sales volumes, the two latter which are what literature primarily have focused on when exploring the influence of reviews, the purpose of this thesis will be to explore the influence of (aggregated) ratings on these variables, while still including that of purchase intentions as previous research has shown that these are linked (Oliver, 1980). We have already defined these four variables as perceived consumer experiences, which therefore will be the formulation in our research question. The choice of aggregated ratings over individual reviews was, as will also be methodologically explained in **section 3.7**, simply because a manipulated number-based rating was easier to control for unintended influences that a fabricated individual review might have. Furthermore, as research has shown that reviews can have an influence on purchase intentions in the pre-experience setting and reconstructive

memory processes regarding advertising information may occur in the post-experience setting, we argue that this is translatable to the case of aggregate ratings in both temporal settings, which is the reason for including this dimension into our research question, that follows below.

Research question: *Can perceived consumer experiences be influenced by ratings both in a pre- and post-experience setting?*

Important to note here is that we intended to research if perceived consumer experiences could be influenced by ratings in two temporal settings, i.e. we are not intending to compare in which temporal setting the perceived consumer experiences can be influenced the most, neither did we intend to measure the extent of the influence. Rather, we allocate our efforts towards establishing support for that the perceived consumer experiences indeed can be influenced in both the pre- and the post-experience setting, separately.

2.6. Hypotheses

The research question was divided into hypotheses according to how we have defined "perceived consumer experiences", i.e. one key indicator of change in perceived consumer experience per hypothesis, as well as one temporal setting. That means that each key indicator is present in two hypotheses, one for the pre-experience setting and one for the post-experience setting. With four key indicators of change in perceived consumer experience and two temporal settings, this leaves us with eight hypotheses, i.e. four a- and b-hypotheses, as shown in **Table 1** below. As previous literature has shown that reviews influence evaluation of an experience in the direction of the review, we assumed this to be the case also for the (aggregated) ratings that we aimed to test our hypotheses with.

Table 1. Summary of hypotheses

	Hypothesis
H1a	Receiving a high rating will generate higher ***fulfillment* of expectations** compared to receiving a low rating in a pre experience setting.
H1b	Receiving a high rating will generate higher ***fulfillment* of expectations** compared to receiving a low rating in a post experience setting.
H2a	Receiving a high rating will generate higher ***rating*** compared to receiving a low rating in a pre experience setting.
H2b	Receiving a high rating will generate higher ***rating*** compared to receiving a low rating in a post experience setting.
H3a	Receiving a high rating will generate higher ***satisfaction*** compared to receiving a low rating in a pre experience setting.
H3b	Receiving a high rating will generate higher ***satisfaction*** compared to receiving a low rating in a post experience setting.
H4a	Receiving a high rating will generate higher ***purchase intentions*** compared to receiving a low rating in a pre experience setting.
H4b	Receiving a high rating will generate higher ***purchase intentions*** compared to receiving a low rating in a post experience setting.

2.7. Theoretical model

In **Figure 1** and **Figure 2**, we have attempted to draw the hypothesized influence of ratings on the constituents of perceived consumer experiences for the pre- and post-experience respectively. **Figure 1** shows how consumers are exposed to a rating *before* going through an experience, which according to our theoretical framework entails that a primacy effect occurs before the experience is realized. The consumer therefore goes through the experience with the rating in the back of their mind, which should lead to them evaluating the fulfillment of expectations, rating, satisfaction and purchase intentions higher when having received a high rating than consumers who had received a low rating.

Figure 1. The pre-experience setting, where consumers are primed with a rating, which according to our theoretical framework should shift the evaluation of their experience in the direction of the rating.
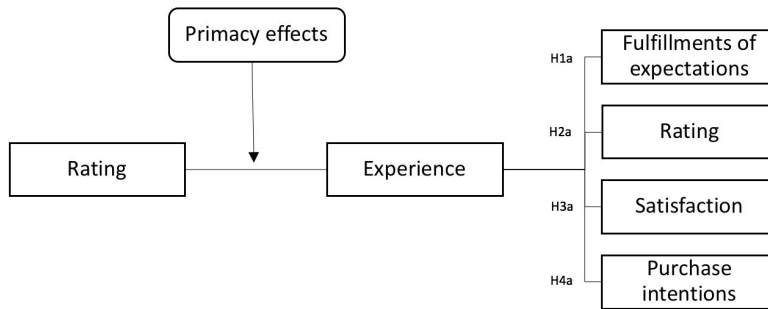
In **Figure 2**, we visualize the influence of ratings received *after* the experience on the variable of perceived consumer experiences. Consumers first go through the experience unaware of what rating it has received by others (true or false) and then receives a rating. Given that the rating received is different from what the consumer would rate the experience him- or herself, the consumer now experiences inconsistent thoughts within their minds. According to consistency theory that we have included into our theoretical framework, this means that the consumer is faced with cognitive dissonance that they will try to reduce. This will according to our theoretical framework be realized through the consumer shifting their evaluation of the experience in the direction of the rating, as the opposite cannot happen, i.e. the received rating cannot be shifted towards the own evaluation.

Figure 2. The post-experience setting, where consumers receiving the rating *after* their experience are left with cognitive dissonance, which according to our theoretical framework should shift the evaluation of their experience in the direction of the rating.
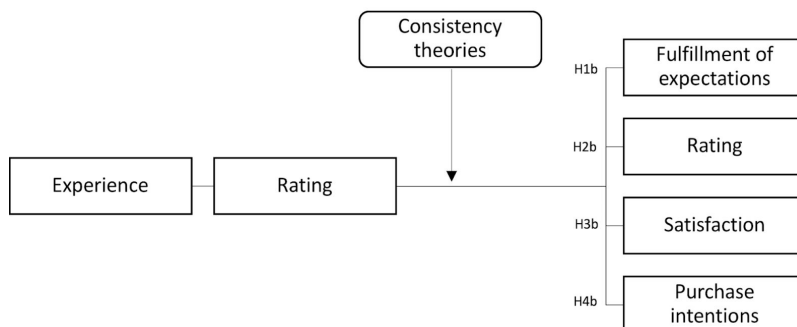
3. Methodology - Study 1

In this section, we aim to motivate the choices we have made regarding method and research design with respect to the chosen research question.

3.1. Methodological/Scientific Approach

The purpose of this study was to answer to the research question: "Can ratings influence the perceived consumer experience in both a pre- and post-experience setting?". Coming from a positivist standpoint, as we assumed empirics would mirror the derived theoretical framework, a quantitative study was motivated because it allows us to gain nomothetic knowledge which predicts reality (Bell & Thorpe, 2013). As we based our hypotheses on existing theory, with a priori research question and tested the hypotheses in an authentic setting, this study has adopted a deductive research approach, (Bryman & Bell, 2011).

3.2. Experimental research design

As the research question: "Can ratings influence the perceived consumer experience in both a pre- and post-experience setting?" implies an assumption of a causal relationship between ratings and the change in perceived consumer experience, we argued that an experimental research design was motivated as this would allow us to manipulate the independent variable, i.e. the received rating in the stimulus. An experiment can be defined as: "individuals being randomly assigned to different groups, which receive different manipulations - then the reactions from the groups are compared after the manipulation" (Söderlund, 2010), which allows for us as researchers to understand the hypothesized causality between the independent variable (ratings) and the dependant variables that we have defined as "the perceived consumer experience", in the different treatment groups. We valued the possibility of being able to show this causality through an experimental research design higher than the possible disadvantages of such a research design[1].

---

[1] According to Bryman & Bell (2011), the experimental research design has the disadvantage that most independent variables that business researchers would like to study are not possible to manipulate in order to see an effect on the dependent variables, e.g. gender. Some manipulations would imply that the researcher would need to intervene too much in people's lives. There is also a fundamental difference between laboratory versus field experiments, where the latter is to be preferred as it is conducted in real-life situations. Achieving this is however hard and Bryman & Bell (2011) even argue that most "field experiments" within business research are not conducted in such real-life situations, rather, scenarios are employed in real-life environments.

We therefore opted for an experiment design that for the respondent included going through an experience and answering a questionnaire about it afterwards, to be carried out in a public space. The respondents were to be allocated to one out of five groups, i.e. four treatment groups and one control group. The four treatment groups included two groups of respondents that received manipulated ratings prior to the experience, and two groups that received the same manipulated ratings afterwards, aiming to test if ratings have an influence in both settings, as suggested by our theoretical framework and in order to answer our research question that includes this dimension. Both the pre- and post-experience setting included one group receiving a manipulated low rating and one group receiving a manipulated high rating. We argued that if we could show that the influence of ratings is symmetrical from the control group, i.e. a high rating would influence consumers to evaluate their perceived experiences higher, and similarly for low ratings to have the influence of lower evaluations, in both settings, we would with our study be able to support a pattern of influence of ratings that holds true in both settings. This was also the reason for including a control group (no stimulus given), in order to make sure that the non-manipulated average rating would lie in-between that of the low and high rating. By establishing support for such a pattern, we argued that we would be able to answer our research question: "Can ratings influence perceived consumer experiences in both a pre- and post-experience setting?".

3.3. Preparatory work

The choice of an experiment design implied the need to choose a product or service category which offered a product or service that would be feasible to distribute to the participants in the experiment. The selection of a specific product within that product category then motivated a pre-test.

3.3.1. Product/service category selection

The choice of product category for the experiment in Study 1 incorporated several components crucial for the possibility to determine the hypothesized influence of the manipulated ratings. To minimize possible influence of other factors than the stimulus, we decided that a product would be preferable over a service to test the hypotheses, as services could be argued to be more susceptible to fluctuations in e.g. quality. We argued that the

input to the experience, i.e. the product, needed to be of constant quality, while still allowing for the output, i.e. the perceived consumer experience, to differ. To achieve that, we decided to appeal to the consumers' sensory responses, i.e. either by taste, hearing, vision, smell or touch, as Hall (2002) had stated that advertising information enhances the sensory experience, and we argued that this could potentially be translatable for reviews and ratings. We further argued that differences in sensory responses could contribute to respondents evaluating the same product differently, all other factors alike, which would motivate an evaluation questionnaire to be answered by the respondents. For reasons of simplicity of product distribution for us as researchers in a public space, we judged that a taste experience would be preferable over any other sensory experience. The product thus needed to be hygienically possible to distribute to the participants. We judged that the chocolate bar product category would satisfy all of the stated criteria.

3.3.2. Selection of chocolate bar

The choice of a specific chocolate bar however came with further criteria. Firstly, we needed something that was likely to not have been tested by the majority of the prospective respondents, to reduce eventual losses of the sample, as Study 1 incorporated the dimension of proximity in time between the experience and the evaluation of it, i.e. a limited delay between product testing and evaluation. Secondly, as the experiment aimed to test if ratings could influence the evaluation of an experience in both positive and negative direction, we aimed to find a product for which the evaluation of the product experience without treatment would be close to normally distributed around 5-6 on a scale of 1 to 10 (10 being the highest) to allow for effects in both directions for the treatment groups.

We aimed to find a product that fulfilled these criteria, by going for a recently launched, heavily marketed, most preferably a "hyped" product, subject to e.g. product extension or co-branding. This strategy was argued to facilitate the collection of data as more people would want to participate in the experiment if they had heard about the product before, but had not tested it yet. We also argued that, since it was an experiment, potential brand or even product bias effects would be present also in our control groups and thus the brand need not influence our choice of product for the experiment.

We found four products fulfilling our criteria, all within the milk chocolate bar category and containing different kinds of best-selling candy from the companies' respective in-house brand portfolios. These were Cloetta's Plopp Gott & Blandat, Cloetta's Plopp Djungelvrål, Toms Ferrari and Toms Mintstång. Both Malaco Djungelvrål (Malaco is a sub-brand of Cloetta) and Toms Mintstång contain liquorice, which made us exclude these alternatives as we saw a risk in that the preference for liquorice is varied among Swedish consumers which could potentially have influenced the results of the study. We then opted for a pre-test with the aim of helping us choose between Cloetta's Plopp Gott & Blandat and Toms Ferrari.

3.4. Pre-test

In order to determine what chocolate bar to use for the stimuli, respondents were asked to taste both of the two chocolate bars, Cloetta's Plopp Gott & Blandat and Toms Ferrari, and rate both on a 10-point scale (10 being the highest). The reason for using a 10-point scale was simply because this was what was intended to be used also for the stimuli in Study 1 (and Study 2 for that matter). We wanted respondents to be able to self-evaluate their own rating according to the rating received and therefore needed it to be consistent. In turn, using a 7-point Likert scale for the rating question, as for the other questions, we argued to be too different from more common rating standards in non-academic settings, where you more often rate products or experiences on a 5 or 10-point-scale. A 5-point-scale was judged to give too little space for nuances in rating between groups. Respondents were also asked if they had tasted the chocolate bar before (1 = Had tested it before, 2 = Had not tested it before).

3.4.1. Pre-test sample

The pre-test was conducted on the basis of a convenience sample close to the Stockholm grocery store from which we bought the pre-test chocolate bars. Bypassing pedestrians, i.e. any first available data source, was asked to taste the two chocolates. As all 35 respondents tasted them both, there was no assignment to different groups. Minimum respondents for the pre-test was set to 30 as suggested by (Bryman & Bell, 2007) to be able to draw generalizable conclusions about the mean rating and standard deviation from this for each of the two chocolate bars.

3.4.2. Results and analysis of pre-test

As **Table 2** shows, the average rating for Toms Ferrari (5.34) was slightly closer to the center of the scale (5.5) than Cloetta's Plopp Gott & Blandat (6.14), which argued for choosing Toms Ferrari over Cloetta's Plopp Gott & Blandat, given our previously stated criteria. However, the standard deviation was also slightly higher (1.781 compared to 1.734) and we concluded that it was more important for us to have a product with a stable rating. None of the respondents had previously tested either chocolate bar. We argued that the slightly more positive evaluation of Cloetta's Plopp Gott & Blandat still allowed for potential positive effects on the 10-point scale in the treatment groups. Cloetta's Plopp Gott & Blandat was therefore chosen for the taste experience experiment and will hereafter be referred to either as "Plopp Gott & Blandat", the abbreviation "Plopp G&B" or just "the chocolate". **Table 3** and **Table 4** show the distribution of ratings among the 35 respondents.

Table 2. Average rating and standard deviation for both chocolate bars.

	N	Mean	Std. Deviation
Rate - Plopp G&B	35	6.14	1.734
Tasted Plopp G&B before?	35	2.00	.000
Rate - Toms Ferrari	35	5.34	1.781
Tasted Toms Ferrari before?	35	2.00	.000

Table 3. Rate Plopp G&B (1=Worst, 10=Best)

Rating	Frequency	Percent
2	1	2.9
3	1	2.9
4	3	8.6
5	7	20.0
6	8	22.9
7	10	28.6
8	2	5.7
9	1	2.9
10	2	5.7
Total	35	100

Table 4. Rate Toms Ferrari (1=Worst, 10=Best)

Rating	Frequency	Percent
2	4	11.4
3	1	2.9
4	5	14.3
5	8	22.9
6	7	20.0
7	6	17.1
8	4	11.4
Total	35	100

3.5. Establishing a relation with Cloetta

After the specific chocolate bar was chosen for Study 1, we initiated email contact with Cloetta. We offered exposure of their newly marketed chocolate bar, Plopp Gott & Blandat, to the potential experiment participants that we encountered, in exchange for Marketing Director Anna Bartholf sending us the chocolate bars needed for the taste experiment, in commercially marketed size (75 gr à piece).

3.6. Questionnaire

The questionnaire consisted of 18 questions (12 measurement questions, 4 control questions and 2 demographic questions) for the treatment groups and 15 questions (12 measurement questions, 1 control question and 2 demographic questions) for the control group, see **appendix 10.3** for the full questionnaire. Five different versions, one for each group, was printed and respondents had to fill in their answers with a pen on a physical paper. The first page consisted of an introduction where it was stated that the respondents' answers will be used for a master thesis at the Stockholm School of Economics. The introduction also mentioned the amount of questions, estimated time of completion, anonymous participation and short instructions regarding the taste test.

Depending on which group respondents were randomly assigned to, the two different versions of the stimulus was given either before or after the instruction to taste the chocolate. The control group only received instructions of when to taste the chocolate. The questions were divided into different sections with a page break when a new section began. This was followed by four control questions aimed at verifying the manipulation. Lastly, demographic questions regarding age and gender were asked.

Two of the measures consisted of a 3-item scale which later was averaged to an index using Cronbach Alpha to verify the internal consistency. The two other measures were carefully selected single item questions. All of the scale questions were measured on a 7-point scale, expect the rating variable since it made more sense to use a 10-point scale similar to the stimuli, as argued for. The "negative" adjectives, such as "Very Dissatisfied", were anchored on the left endpoint and the "positive" bi-polar antonyms such as "Very Satisfied" were anchored to the right (Söderlund, 2005).

3.6.1. Dependent variables

Fulfillment of expectations

To measure respondents perceived fulfillment of their expectations on the experience, the question "How was the taste experience according to your expectations" was asked. The answer could be specified on a 7-point likert scale with the two antonyms "Much worse than expected" and "Much better than expected" at the endpoints. The measurement was inspired by Richard L. Oliver (1980).

Rating

Respondents were asked to rate the chocolate on the same 10-point scale that were used in the stimuli. The antonyms on each endpoint were "1=worst" and "10=best".

Satisfaction

The experience satisfaction was measured using a 3-item, 7-point semantic differential scale asking respondents to "Evaluate the taste experience". The antonyms used in each endpoint of each item were "Very Displeased/Very Pleased", "Unhappy With/Happy With" and "Very Dissatisfied/Very Satisfied". The question and items were adopted from Jones, Mothersbaugh

& Betty (2000) and were averaged to an index using Cronbach Alpha with an Alpha value of 0.958.

Purchase intentions

Respondents purchase intentions were measured using the question "If you were buying snacks, what is the probability you would consider this chocolate?" measured on a 3-item, 7-point semantic differential scale. The antonyms used in each items endpoints were "Unlikely/Likely", "Very Unprobable/Very Probable" and "Impossible/Possible". The question and items are adopted from Chapman & Aylesworth (1999) and were averaged to an index using Cronbach Alpha with an Alpha value of 0.960.

3.6.2. Control questions

To be able to verify that the respondents had understood the independent variable correctly, four manipulation control questions were asked. Firstly, respondents were asked "What was the rating of the chocolate?" to verify that the respondent had perceived and comprehended the stimulus. Three possible answers were provided: correct, wrong or "I don't know". Secondly, respondents had to indicate whether they perceived the received rating as a generally high or low rating through the question "The rating I chose in the previous question is generally a … rating". Three possible answers were provided: "high", "low" or "I don't know". Thirdly, respondents answered the question "Was the rating credible?" with the two alternatives "yes" or "no". Lastly, to make sure that respondents didn't have any preconceptions from having tasted the chocolate before, the question "Have you tasted this chocolate before?" was asked. Two alternative answers were provided: "yes" or "no".

3.7. Stimuli

As pictorially received false information led to greater false memory creation (Braun-LaTour, LaTour, Pickrell & Loftus, 2004), we opted for stimuli, for each of the four treatment groups respectively, that consisted of a manipulated, star-based rating, i.e. pictorial, also with a picture of the product, an average rating on a 10-point scale and a number of participants in the study. All numbers included were completely fabricated and all stimuli are included in **appendix 10.2**. The use of a manipulated rating from an expert panel was discussed, but as

the effect of online reviews, i.e. eWOM, has been suggested to be nearly identical in significance and importance when the source was user-based rather than expert-based (Amblée & Bui, 2007), we opted for as user-based stimuli as this would not require also the fabrication of an expert panel. Similarly, we saw an advantage in using aggregated ratings as manipulation over individual reviews, as the latter would demand us fabricating more than just a rating score, i.e. review text that potentially would entail unintended influences on the variables tested. As review valence had been shown to have a greater impact on consumers' purchase decisions than review volume (Yang, 2016), we also chose manipulated ratings that were quite far off from each other, 2.4/10 and 9.6/10, respectively, in an attempt to induce the hypothesized influence of the rating to an extent as large as possible. This entailed that we wanted to hedge for credibility of the rating scores being an issue when we opted for such extreme ratings, which was the reason for why we included a credibility confound check into the questionnaire, such as described in **section 3.6.2**, in order to be able to sort out cases that did not perceive the rating to be credible. The control group did not receive any stimulus, i.e. they did not receive the authentic rating either.

3.8. Sampling and assignment

We aimed to have 30 participants for each group as we wanted to run parametric statistical tests rather than non-parametric tests. The taste experiment included a non-convenience sample of 167 respondents, gathered in public spaces around Stockholm and its close surroundings. The sample was thus not random, however the assignment was. Paper questionnaires were put together for the five different groups and then randomly shuffled using a mobile app called "Pretty Random".

3.8.1. Sample demography and cases sorted out

If a respondent did not pass the control questions, their answer was examined to see whether they seemed to have misunderstood the question, if so they were still included and if not, they were omitted away. An example is when a respondent in one of the two high rating treatment groups (received stimulus of 9.6 / 10) had evaluated the majority of the variables very low, i.e. 1 or 2 on scale questions and then answered the correct alternative of "9.6 / 10" on the control question: "What was the rating of the chocolate?", but "Low" (wrong alternative) for

the control question: "The rating I chose in the previous question is generally a … rating".
We argue that such a case would imply that the latter control question was mistakenly
referred back to one's own low ratings on other variables, rather than referring to the
stimulus, as intended. **Table 5** and **Table 6** show some basic demographic characteristics of
the final sample for Study 1, while **Table 7** and **Table 8** present the number of cases sorted
out and the allocation to each treatment/control group, respectively.

Table 5. Gender

Gender	Number	Percent
Female	83	49.7
Male	83	49.7
Other	1	0.6
Total	167	100

Table 6. Age

Age	Min	Max	Mean	Std. Dev.
	18	74	33.02	12.106

Table 7. Quality control

Quality Control of data	
Initial Sample	183
Cases sorted out	16
Final Sample	167

Table 8. Assignment

Group	N
Control	40
Before Low	30
Before High	32
After Low	33
After High	32
Total	167

3.9. Analytical tools and statistical tests

IBM SPSS Statistics version 21 was used to analyse the collected data for Study 1 and the following tests were used:

- Cronbach Alpha
- One-way ANOVA
- Non-parametric K independent sample (Kruskal-Wallis)

The hypotheses were accepted at a level of significance of 95% (Fisher 1992). Significance levels are presented in each test as stated below.

* Significant at $p < .05$
** Significant at $p < .01$
*** Significant at $p < .001$

3.10. Data quality

The data quality in terms of reliability and validity for both studies will be jointly evaluated in **section 6.9**, in order for us to first be able to explain the methodological choices we have made also for Study 2.

4. Results & Analysis - Study 1

Findings for Study 1 are presented and followed by our concluding remarks that argues for an additional study. The complete data output from spss can be found in appendix.

Table 9. Recap of hypotheses tested in Study 1

Hypothesis	
H1a	Receiving a high rating will generate higher ***fulfillment* of expectations** compared to receiving a low rating in a pre experience setting.
H1b	Receiving a high rating will generate higher ***fulfillment* of expectations** compared to receiving a low rating in a post experience setting.
H2a	Receiving a high rating will generate higher *rating* compared to receiving a low rating in a pre experience setting.
H2b	Receiving a high rating will generate higher *rating* compared to receiving a low rating in a post experience setting.
H3a	Receiving a high rating will generate higher *satisfaction* compared to receiving a low rating in a pre experience setting.
H3b	Receiving a high rating will generate higher *satisfaction* compared to receiving a low rating in a post experience setting.
H4a	Receiving a high rating will generate higher *purchase intentions* compared to receiving a low rating in a pre experience setting.
H4b	Receiving a high rating will generate higher *purchase intentions* compared to receiving a low rating in a post experience setting.

4.1. One-Way ANOVA

The results below are from testing the whole population.

Fulfillment of expectations

Table 10. Fulfillment of expectations - Before

Fulfillment of expectations	Before Low		Control		Before High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H1a	4.80	1.35	-	-	5.22	1.36	0.42	.827
	4.80	1.35	4.13	1.38	-	-	0.67	.370
	-	-	4.13	1.38	5.22	1.36	1.09	.022*

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 11. Fulfillment of expectations - After

Fulfillment of expectations	After Low		Control		After High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H1b	4.61	1.39	-	-	5.34	1.21	0.73	.305
	4.61	1.39	4.13	1.38	-	-	0.48	.680
	-	-	4.13	1.38	5.34	1.21	1.21	.007**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called fulfillment of expectations, when receiving the stimulus before the experience, the mean difference between the low rating (mean = 4.80) and high rating (mean = 5.22) is 0.42. The test reveals no statistical difference and H1a is therefore rejected. The mean for the control group is lower than both the low and high group and a statistical difference is found between the control group and the high group.

When receiving the stimulus after the experience, the mean difference between the low group (mean = 4.61) and high group (mean = 5.34) is 0.73. The test reveals no statistical difference and H1b is therefore rejected. The mean for the control group is lower than both the low and high group and a statistical difference is found between the control group and the high group.

Rating

Table 12. Rating - Before

Rating	Before Low		Control		Before High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H2a	6.03	2.24	-	-	6.59	2.06	0.56	.869
	6.03	2.24	5.68	1.86	-	-	0.35	.966
	-	-	5.68	1.86	6.59	2.06	0.91	.428

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 13. Rating - After

Rating	After Low		Control		After High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H2b	5.76	1.79	-	-	6.75	1.93	0.99	.394
	5.76	1.79	5.68	1.86	-	-	0.08	1.000
	-	-	5.68	1.86	6.75	1.93	1.07	.264

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called rating, when receiving the stimulus before the experience, the mean difference between the low rating (mean = 6.03) and high rating (mean = 6.59) is 0.56. The test reveals no statistical difference and H2a is therefore rejected. The mean for the control group is lower than both the low and high group but no statistical difference is found between the groups.

When receiving the stimulus after the experience, the mean difference between the low group (mean = 5.76) and high group (mean = 6.75) is 0.99. The test reveals no statistical difference and H2b is therefore rejected. The mean for the control group is lower than both the low and high group but no statistical difference is found between the groups.

Satisfaction

Table 14. Satisfaction - Before

Satisfaction	Before Low		Control		Before High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H3a	4.76	1.47	-	-	5.00	1.40	0.24	.971
	4.76	1.47	4.38	1.26	-	-	0.38	.855
	-	-	4.38	1.26	5.00	1.40	0.62	.436

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 15. Satisfaction - After

Satisfaction	After Low		Control		After High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H3b	4.41	1.35	-	-	5.20	1.19	0.79	.223
	4.41	1.35	4.38	1.26	-	-	0.03	1.000
	-	-	4.38	1.26	5.20	1.19	0.82	.152

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called satisfaction, when receiving the stimulus before the experience, the mean difference between the low rating (mean = 4.76) and high rating (mean = 5.00) is 0.24. The test reveals no statistical difference and H3a is therefore rejected. The mean for the control group is lower than both the low and high group but no statistical difference is found between the groups.

When receiving the stimulus after the experience, the mean difference between the low group (mean = 4.41) and high group (mean = 5.20) is 0.79. The test reveals no statistical difference and H3b is therefore rejected. The mean for the control group is lower than both the low and high group but no statistical difference is found between the groups.

Purchase intentions

Table 16. Purchase intentions - Before

WTB	Before Low		Control		Before High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H4a	3.43	1.88	-	-	3.84	1.84	0.41	.937
	3.43	1.88	3.23	1.62	-	-	0.20	.995
	-	-	3.23	1.62	3.84	1.84	0.61	.725

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 17. Purchase intentions - After

Satisfaction	After Low		Control		After High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H4b	3.06	1.76	-	-	3.82	1.90	0.76	.570
	3.06	1.76	3.23	1.62	-	-	0.17	.997
	-	-	3.23	1.62	3.82	1.90	0.59	.750

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable purchase intentions, when receiving the stimulus before the experience, the mean difference between the low rating (mean = 3.43) and high rating (mean = 3.84) is 0.41. The test reveals no statistical difference and H4a is therefore rejected. The mean for the control group is lower than both the low and high group but no statistical difference is found between the groups.

When receiving the stimulus after the experience, the mean difference between the low group (mean = 3.06) and high group (mean = 3.82) is 0.79. The test reveals no statistical difference and H4b is therefore rejected. The mean for the control group is between the low and high group but no statistical difference is found between the groups.

Result of credibility confound check

None of the hypotheses could be supported in this test, therefore further effort and investigation of the data, and reason behind it, was taken. Through our control question, "was the rating credible?", it was discovered that a some of the respondents did not believe the stimulus to be credible due to reasons discussed in the limitations section 7.3. For the

following tests these respondents, accumulated to 59 cases, were sorted out of the data. However, the number of respondents in each group did not reach our preferred minimum number of 30 and therefore a non-parametric test has been used to further investigate if the hypotheses could be supported.

4.2. Non-parametric K independent sample - Kruskal-Wallis

The results below are only the respondents which perceived the stimulus to be credible.

Table 18. Assignment

Group	*N*
Control	40
Before Low	13
Before High	17
After Low	22
After High	16
Total	108

Fulfillment of expectations

Table 19. Fulfillment of expectations - Before

Fulfillment of expectations	Before Low	Control	Before High	Kruskal-Wallis H	Adj. Sig.
H1a	39.69	-	81.21	41.51	.002**
	39.69	43.54	-	3.85	1.000
	-	43.54	81.21	37.67	.000***

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 20. Fulfillment of expectations - After

Fulfillment of expectations	After Low	Control	After High	Kruskal-Wallis H	Adj. Sig.
H1b	45.14	-	78.44	33.30	.010*
	45.14	43.54	-	1.60	1.000
	-	43.54	78.44	34.90	.001**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called fulfillment of expectations, when receiving the stimulus before the experience, the mean rank difference between the low rating (mean rank = 39.69) and high rating (mean rank = 81.21) is 41.52. The test reveals a statistical difference and H1a is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group and a statistical difference is found between the control group and the high group.

When receiving the stimulus after the experience, the mean rank difference between the low group (mean rank = 45.14) and high group (mean rank = 78.44) is 33.30. The test reveals a statistical difference and H1b is therefore supported on a 1% significance level. The mean rank for the control group is lower than both the low and high group and a statistical difference is found between the control group and the high group.

Rating

Table 21. Rating - Before

Rating	Before Low	Control	Before High	Kruskal-Wallis H	Adj. Sig.
H2a	34.54	-	81.53	46.99	.000***
	34.54	48.80	-	14.26	1.000
	-	48.80	81.53	32.73	.003**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 22. Rating - After

Rating	After Low	Control	After High	Kruskal-Wallis H	Adj. Sig.
H2b	37.14	-	80.12	42.99	.000***
	37.14	48.80	-	11.66	1.000
	-	48.80	80.12	31.32	.006**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called rating, when receiving the stimulus before the experience, the mean rank difference between the low rating (mean rank = 34.54) and high rating (mean rank = 81.53) is 46.99. The test reveals a statistical difference and H2a is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group and a statistical difference is found between the control group and the high group.

When receiving the stimulus after the experience, the mean rank difference between the low group (mean rank = 37.14) and high group (mean rank = 80.12) is 42.99. The test reveals a statistical difference and H2b is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group and a statistical difference is found between the control group and the high group.

Satisfaction

Table 23. Satisfaction - Before

Satisfaction	Before Low	Control	Before High	Kruskal-Wallis H	Adj. Sig.
H3a	35.04	-	78.85	43.81	.001**
	35.04	48.31	-	13.27	1.000
	-	48.31	78.85	30.54	.007**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 24. Satisfaction - After

Satisfaction	After Low	Control	After High	Kruskal-Wallis H	Adj. Sig.
H3b	38.89	-	81.38	42.49	.000***
	38.89	48.31	-	9.43	1.000
	-	48.31	81.38	33.06	.003**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called satisfaction, when receiving the stimulus before the experience, the mean rank difference between the low rating (mean rank = 35.04) and high rating (mean rank = 78.85) is 43.81. The test reveals a statistical difference and H3a is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group and a statistical difference is found between the control group and the high group.

When receiving the stimulus after the experience, the mean rank difference between the low group (mean rank = 38.89) and high group (mean rank = 81.38) is 42.49. The test reveals a statistical difference and H3b is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group and a statistical difference is found between the control group and the high group.

Purchase intentions

Table 25. Purchase intentions - Before

WTB	Before Low	Control	Before High	Kruskal-Wallis H	Adj. Sig.
H4a	32.15	-	78.24	46.08	.001**
	32.15	53.76	-	21.61	.301
	-	53.76	78.24	24.47	.068

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

Table 26. Purchase intentions - After

WTB	After Low	Control	After High	Kruskal-Wallis H	Adj. Sig.
H4b	36.16	-	74.50	38.34	.002**
	36.16	53.76	-	17.60	.336
	-	53.76	74.50	20.74	.247

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable purchase intentions, when receiving the stimulus before the experience, the mean rank difference between the low rating (mean rank = 32.15) and high rating (mean rank = 78.24) is 46.08. The test reveals a statistical difference and H4a is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group but no statistical difference is found between the groups.

When receiving the stimulus after the experience, the mean rank difference between the low group (mean rank = 36.16) and high group (mean rank = 74.50) is 38.34. The test reveals a statistical difference and H4b is therefore supported on a 1% significance level. The mean rank for the control group is between the low and high group but no statistical difference is found between the groups.

4.3. Summary of results - Study 1

Table 27. Support of hypotheses

	Hypothesis	One-way Anova (All respondents)	Non-Parametric (Only credible)
H1a	Receiving a high rating will generate higher *fulfillment of expectations* compared to receiving a low rating in a pre experience setting.	Not Supported	Supported
H1b	Receiving a high rating will generate higher *fulfillment of expectations* compared to receiving a low rating in a post experience setting.	Not Supported	Supported
H2a	Receiving a high rating will generate higher *rating* compared to receiving a low rating in a pre experience setting.	Not Supported	Supported
H2b	Receiving a high rating will generate higher *rating* compared to receiving a low rating in a post experience setting.	Not Supported	Supported
H3a	Receiving a high rating will generate higher *satisfaction* compared to receiving a low rating in a pre experience setting.	Not Supported	Supported
H3b	Receiving a high rating will generate higher *satisfaction* compared to receiving a low rating in a post experience setting.	Not Supported	Supported
H4a	Receiving a high rating will generate higher *purchase intentions* compared to receiving a low rating in a pre experience setting.	Not Supported	Supported
H4b	Receiving a high rating will generate higher *purchase intentions* compared to receiving a low rating in a post experience setting.	Not Supported	Supported

4.4. Concluding remarks - Study 1

The confound check regarding credibility of stimuli showed that a high share of the respondents did not perceive the rating to be credible, which entailed that the empirics did not support the hypotheses using a One-way Anova test. However, excluding the respondents that did not find the received stimulus to be credible allowed us to see the hypothesized influence using a non-parametric test. The results of this supported all hypotheses. As the exclusion of respondents led to small samples of each treatment group which are generally not considered reliable enough to be able to generalize to a bigger population (Mundry & Fischer, 1998;

Conover & Iman, 1981), we opted for an additional study aimed towards replicating the results of Study 1 to further strengthen our findings through a One-way Anova test and with a larger sample. Collecting more cases for Study 1 was discussed, but we argued that allocating our time to an additional experiment in a different context would instead allow us to study if the hypotheses supported by the non-parametric tests would be supported also by the One-way Anova test in a different context. We argued that this method would show even greater support for our hypotheses than if we would have just collected more cases for Study 1.

5. Methodology - Study 2

This section aims to describe the methodology for Study 2, what is different from Study 1 and our arguments for making these alterations. As the motivation for Study 2 was to strengthen our findings from Study 1, much of our methodological approach is the same for Study 2 as for Study 1. Unless stated otherwise, the reader may assume that the experiment has been executed in a very similar fashion.

5.1. Product category selection

The product category selection preceded in this study the specificities of the experimental research design, due to our deductive approach. The chocolate bar tested in the taste experiment can, with respect to our chosen theoretical framework, be considered both a search good, i.e. a snack or a type of sweet, and an experience good, i.e. a taste experience, depending on how much emphasis you put on the hedonistic aspect of it being consumed. As existing theory, as explained previously, have concluded that the influence of reviews on consumer evaluation differ depending on if the product contained more or less hedonic versus utilitarian (ambiguous/unambiguous) properties as well as the extent to which positive reviews have an influence on search and experience goods, respectively, differs, we aimed for a product category different to that of chocolate bars. With a product that with more certainty can be considered an experience good - a film experience in a theatre setting - we argued that we could complement and strengthen our findings in the main study. Also, as we aimed to study the difference in impact between the low and high rating, an experience good would be the ultimate choice as Hao, Ye, Li & Cheng (2010) had shown that the difference between a positive and negative review in impact on evaluation was higher for experience goods than for search goods, and we argued that this would be translatable to the case of ratings, as defined as aggregated reviews.

5.2. Establishing a relation with Filmstaden

Thanks to our fellow student within the master program in Business and Management, Zacharias Lindqvist Hansson, that had interned at the headquarters of the Swedish cinema

chain Filmstaden, we could make direct contact with Filmstaden's Head of Analysis, Robin Fischer. He was involved with the experiment, to some degree, from planning to execution.

5.3. Experimental research design

The experiment was designed much like the taste experiment in Study 1, with an experience after which the respondents needed to answer a questionnaire about their perceived experience. For reasons of transparency towards Filmstaden's customers, we were not able to send our stimuli via email to the filmgoers that had not yet seen the film, followed by an email questionnaire after the movie, as this would implicate that some filmgoers might not get the information that the pre-experience ratings had been manipulated. This led to our exclusion of the two pre-experience treatment groups for this experiment, Study 2, and meant that we could only test our b-hypotheses in Study 2, i.e. the after-setting hypotheses. We were aware of this trade-off when opting for Study 2 over expanding the sample for Study 1, but argued that two studies in different contexts, whereof one would support all of our hypotheses through non-parametric tests and both would support our b-hypotheses through a non-parametric test and a One-way Anova test, respectively, would be stronger than only one study in one context.

As respondents most likely had went home from the theatre and a day or two had passed since they had seen the film before answering the questionnaire, this study also incorporated a dimension of *delay* between the product-testing and the stimulus information, as suggested by Srull & Wyer (1980) to preclude retrieval efforts of information consistent with the stimulus, and thus diminishing the extent to which respondents would shift the evaluation of their experience in the direction of the review (Hoch & Ha, 1986). It would therefore not be unlikely to see less prominent support for Study 2 than for Study 1 that did not incorporate this delay, which we argued would only be an additional strengthening factor for that our findings would hold over different contexts.

As the pre-experience treatment groups needed to be excluded, this study therefore included three groups rather than five, i.e. the two post-experience treatment groups, low and high, as well as a control group. In this study also, the purpose of the control group was to see if its

average perceived consumer experience followed the pattern where it was in-between that of the low and high groups' evaluation of their perceived experiences, as was the case with both the pre- and post-experience settings for the taste experiment in Study 1, except for the variable fulfillment of expectations in the post experience setting.

We argued that, if we could show that the same pattern, that we supported with non-parametric tests in both a pre- and post-experience setting in Study 1, holds also for the post-experience setting in Study 2, we would have found even stronger support for all of our b-hypotheses, **Table 28.** With strong support for the b-hypotheses and thus for ratings influence in the post-experience setting, we then argued that if the same pattern for the pre-experience setting existed in Study 1 (using non-parametric tests) as in the post-experience setting in Study 1 and Study 2 (using non-parametric tests and the One-way Anova test, respectively) we argued that the combination of the three findings would be strong enough to be able to answer our research question in full. Important to state is that we are not looking to support which point in time, i.e. pre- or post-experience, in which ratings have the greatest influence, only that they do have an influence in both settings. We argued that the pattern of average evaluations in the low and high groups for both settings, and with the control group in-between, would allow for us to support that ratings do indeed have an influence in both a pre- and post-experience setting, even if we cannot determine when they have the greatest influence with this particular study.

Table 28. Recap of our b-hypotheses.

Hypothesis	
H1b	Receiving a high rating will generate higher ***fulfillment* of expectations** compared to receiving a low rating in a post experience setting.
H2b	Receiving a high rating will generate higher ***rating*** compared to receiving a low rating in a post experience setting.
H3b	Receiving a high rating will generate higher ***satisfaction*** compared to receiving a low rating in a post experience setting.
H4b	Receiving a high rating will generate higher ***purchase intentions*** compared to receiving a low rating in a post experience setting.

5.4. Film selection

The choice of film was motivated by a criterion of its rating not being too easily judged or perceived, e.g. visible on the commercial ad for the film. Neither should it have been running in the theatres for too long before the experiment, as this would entail a risk of more people having seen ratings or heard other people talk about it, which might have had an influence on the results of our study. The film Halloween was suggested by the Head of Analysis at Filmstaden and was chosen as it fulfilled these criteria. It premiered on October 19, 2018, and the experiment was carried out in full between November 13 and 14, 2018.

5.5. Questionnaire

The questionnaire for the film experiment was based on the taste experiment questionnaire and are to be found in **appendix 10.6**. To the extent it was possible, the same questions was included and altered as little as possible to suit the evaluation of a film experience, as we aimed to link the results of both studies. The first question was a added screening question: "Have you seen the movie Halloween in the theatre?". Other than that, four of the 18 questions received alterations worth mentioning. The taste specific question for the taste experience: "To what extent does the chocolate bar taste Gott & Blandat?" (Very little/Very much) corresponds to the genre-specific film experience question: "I think the film was…" (Not at all scary/Very scary). The measurement for purchase intentions was changed from "If you were buying snacks, what is the probability you would consider this chocolate?" in the taste experiment to "How would you feel about a sequel to this movie" in the film experience. The manipulation control question "Have you tasted this chocolate before?" was changed to "Have you seen other ratings for this film?". Lastly, a control question "What was the name of the film you saw?" was added in complement to the first screening question with one correct answer, one faulty answer and the option to check "I don't know".

Another difference from Study 1 was that the questionnaire was distributed digitally using the online questionnaire tool Qualtrics. An advantage with this was that respondents were forced to answer all the questions in order to proceed. The order of the questions were the same as Study 1.

5.5.1. Dependent variables

Fulfillment of expectations

To measure respondents fulfillment of expectations the question "How was the movie experience according to your expectations" was asked. The answer could be specified on a 7-point Likert scale with the two antonyms "Much worse than expected" and "Much better than expected" at the endpoints. The measurement was inspired by Study 1 and Richard L. Oliver (1980).

Rating

Respondents were asked to rate the movie on the same 10-point scale that were used in the stimuli. The antonyms on each endpoint were "1=worst" and "10=best".

Satisfaction

The experience satisfaction was measured using a 3-item, 7-point semantic differential scale asking respondents to "Evaluate the movie experience". The antonyms used at each endpoint of each item were "Very Displeased/Very Pleased", "Unhappy With/Happy With" and "Very Dissatisfied/Very Satisfied". The question and items are adopted from study 1 and Jones, Mothersbaugh & Betty (2000) and were averaged to an index using Cronbach Alpha with an Alpha value of 0.923.

Purchase intentions

Respondents purchase intentions were measured using the question "How would you feel about a sequel to this movie?" measured on a 3-item, 7-point semantic differential scale. The antonyms used at each items endpoints were "Don´t wan't to see/Wan't to see", "Indifferent/Curious" and "Uninterested/Interested". The question and items was inspired by the taste experiment but had to be adopted to the situation with the recommendation from Micael Dahlen, tutoring Professor. It can however be argued that the question is more phrased as repurchase intentions. The three items were averaged to an index using Cronbach Alpha with an Alpha value of 0.959.

5.6. Stimuli

The stimuli for the film experiment were made upon the basis of the chocolate stimuli and are to be found in **appendix 10.4**. However, as ratings are more frequently given on consumers' own accord and thus more available to the public for film, than for chocolate, we perceived there to be a credibility risk if including too high or too low manipulated ratings in the stimuli. Also, reviews with extreme ratings have, as previously explained, been shown to be perceived as less helpful than reviews with moderate rating for experience goods (Mudambi & Schuff, 2010), e.g. films. After discussions with the Head of Analysis at Filmstaden, Robin Fischer, we judged that manipulated ratings deviating by around 1.5 from the average score would allow us to see the hypothesized influence, while still being credible. To estimate the average score for the film in question, we used imdb.com, which in Sweden is a commonly used online rating platform for film.

5.7. Sampling and assignment

Filmstaden kindly accommodated our request to send our questionnaire with the film experiment to their customers electronically. Advantageously, this meant that the sample was not geographically restricted to where we as researchers' live and work, i.e. Stockholm, as would be the case if the experiment was to be conducted e.g. in one of Filmstaden's theatres.

The European Union general data protection regulation (GDPR; effective from May 25, 2018) posed obstacles regarding email marketing consent for Filmstaden to be able to send our survey to any filmgoer of theirs, as only Filmstaden's club members had agreed to such email sendouts through their membership terms. For reasons of time constraint, we needed to incentivize people to participate in the experiment, and the questionnaire was thus sent to Filmstaden's film panel, through which respondents were compensated by Filmstaden for participating in the film experiment.

Random assignment was accomplished through the Qualtrics tool determining which treatment group each respondent were directed to when clicking the email questionnaire link.

5.7.1. Sample demography and cases sorted out

The fact that the respondents were compensated by Filmstaden to participate in our study entailed a high number of respondents clicking the email questionnaire link. It also explains the large number of respondents removed from the initial sample, both as some respondents failed the control questions, but also as many, based on the time taken to complete the questionnaire, seem to have been just randomly clicking their way through the questionnaire, focusing more on the monetary compensation and finishing quickly than providing accurate data. However, the data was thoroughly examined and these cases were excluded from the sample. **Table 29** and **Table 30** show some basic demographic characteristics of the final sample, while **Table 31** and **Table 32** show how many cases were sorted out from the initial sample and the allocation to the each treatment/control group, respectively.

Table 29. Gender

Gender	Number	Percent
Female	75	44.1
Male	95	55.9
Total	170	100

Table 30. Age

Age	Min	Max	Mean	Std. Dev.
	18	64	35.60	10.750

Table 31. Quality control

Quality Control of data	
Initial Sample (clicking the email questionnaire link)	1855
Cases sorted out in initial screening (what film they had seen)	1469
Survey-taking sample	386
Cases sorted out	216
Final Sample	170

Table 32. Assignment

Group	N
Control	61
Low	62
High	47
Total	170

5.8. Analytical tools and statistical tests

IBM SPSS Statistics version 21 was used to analyse the collected data for Study 2 and the following tests were used:

- Cronbach Alpha
- One-way ANOVA

The hypotheses were accepted at a level of significance of 95% (Fisher 1992). Significance levels are presented in each test as stated below.

* Significant at $p < .05$

** Significant at $p < .01$

*** Significant at $p < .001$

5.9. Data quality (for both Study 1 and 2)

For the purpose of the two studies in this thesis and quantitative research generally, it is of high importance to reflect over and conduct the research so that the topic researched is accurately measured and provide a precise description of reality (Bryman & Bell, 2015). The reliability and validity of the two studies are evaluated below.

5.9.1. Reliability

The consistency of measurements and if the results of the study is replicable through other studies over time is referred to as the reliability of the study according to Bryman & Bell (2015).

The thesis consists of two main studies, studying the same research question in two different contexts in a very similar way. The results of these studies is in line with each other, indicating replicatibility both over time and contexts. The respondents in both studies were not a random sample, in Study 1 the sample consisted of random people in the streets of Stockholm, however the assignment to the different treatment groups were random. For Study 2 the sample consisted of people who had seen a specific movie and was a member of Filmstaden's film panel, however, the assignments to the different treatment groups were random. Both of the samples that are of non-convenience type, can be argued to be reliable as samples that represent a wider population. Further, the purpose of the study was not revealed to the respondents before answering the questionnaire, which lowers the risk of unintended priming effects.

For both studies, manipulation checks in terms of control question was asked to ensure that the respondents had understood the stimulus correctly, hence minimizing the risk of bias in both studies. Questions asked in the questionnaire for both studies were, in the absolute majority, adopted from established scientific articles. To secure a higher level of internal reliability some measures were also multi-item measurements (Söderlund, 2005), which was later averaged into indexes using Cronbach's Alpha. The alpha values was only accepted when exceeding 0.7 and the values ranged from 0.923 to 0.962 indicating a high internal consistency which increases the reliability (Bryman & Bell, 2015).

Evaluating reliability

Considering the actions taken above to ensure the studies being reliable, there is little reason to believe the reliability would be below a satisfactory level.

5.9.2. Validity

Validity refers to what extent the data collected actually measured what was intended. Validity is divided into *internal*, which aims to describe if causal relationships can be made between the dependent and independent variables, and *external*, which describes to what degree the results can be generalized (Söderlund 2005).

Internal validity

The internal validity for the two studies is to what extent casual arguments can be made regarding how different ratings (independent variable) affect fulfillment of expectations, rating, satisfaction and purchase intentions (dependent variables) (Söderlund 2010). Both studies had a control group and respondents were randomly assigned to the different groups, this addresses crucial threats to the internal validity (Söderlund 2010). Multiple manipulation check questions was used to ensure that the respondents had both understood the stimulus correctly and respondents were sorted out of the population if they had gone through the experience beforehand. To increase the internal validity further, questions and measurements used were adopted from established scientific literature and multi-item measurements were used (Bryman and Bell, 2007).

External validity

The external validity regards how causal arguments made in the studies can be applied or generalized to other situations or contexts (Malhotra & Birks, 2007). This thesis covers two studies executed in a very similar fashion in two different contexts with results pointing in the same direction, indicating a high external validity. Both of the experiments were conducted in non-laboratory settings with non-convenience samples. For the taste experiment, the preferred number of respondents could not be reached, however this was compensated by conducting another study, the film experiment, where the preferred number of respondents in each treatment group was reached. The population for the taste experiment was aged between 18-74 (mean = 33.02) and the gender distribution was even (Female = 49.6%, Male = 50%). It was collected in public spaces around Stockholm and can be argued to represent a population similar to the whole of Sweden. However, to further strengthen this, the population in the film experiment was nationwide and strong arguments can certainly be

made that this is a valid representation for the Swedish people as a whole population. The age was between 18-64 (mean = 35.60) and the gender distribution was fairly even (Female = 44.1%, Male = 55.9%). The film experiment, can be argued to be a true experiment were the respondents were studied in a real-life setting. The fact that this experiment studied natural occurring phenomenon, except that the stimulus was constructed and sent afterwards, indicates a high ecological validity (Bryman & Bell, 2015).

Evaluating validity

Considering the actions taken and arguments made regarding the validity, both internal and external, for both studies, there is little reason to believe that the validity should not reach a satisfactory level.

6. Results & Analysis - Study 2

Findings regarding each of four variables are presented, after which our concluding remarks around these will follow. A more thorough discussion of our findings will follow in the next chapter and the complete data output from spss can be found in appendix.

Table 33. Recap of hypotheses tested in Study 2

Hypothesis	
H1b	Receiving a high rating will generate higher ***fulfillment of expectations*** compared to receiving a low rating in a post experience setting.
H2b	Receiving a high rating will generate higher ***rating*** compared to receiving a low rating in a post experience setting.
H3b	Receiving a high rating will generate higher ***satisfaction*** compared to receiving a low rating in a post experience setting.
H4b	Receiving a high rating will generate higher ***purchase intentions*** compared to receiving a low rating in a post experience setting.

6.1. One-Way ANOVA

Table 34. Fulfillment of expectations

Fulfillment of expectations	Low		Control		High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H1b	4.27	1.23	-	-	5.34	1.03	1.07	.000***
	4.27	1.23	4.77	1.20	-	-	0.50	.065
	-	-	4.77	1.20	5.34	1.03	0.57	.045*

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called fulfillment of expectations, the mean difference between the low rating (mean = 4.27) and high rating (mean = 5.32) is 1.07. The test reveals a statistical difference and H1b is therefore supported on a 1% significance level. The mean for the

control group is between the low and high group and a statistical difference is found between the control group and the high group.

Table 35. Rating

Rating	Low		Control		High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H2b	5.47	1.65	-	-	7.36	1.36	1.89	.000***
	5.47	1.65	6.41	1.69	-	-	0.94	.005**
	-	-	6.41	1.69	7.36	1.36	0.95	.010*

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called rating, the mean difference between the low rating (mean = 5.47) and high rating (mean = 7.36) is 1.89. The test reveals a statistical difference and H2b is therefore supported on a 1% significance level. The mean for the control group is between the low and high group and a statistical difference is found between both the control group and the low group as well as the control group and the high group.

Table 36. Satisfaction

Satisfaction	Low		Control		High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H3b	4.46	1.21	-	-	5.69	0.99	1.23	.000***
	4.46	1.21	4.90	1.10	-	-	0.43	.101
	-	-	4.90	1.10	5.69	0.99	0.79	.002**

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable called satisfaction, the mean difference between the low rating (mean = 4.46) and high rating (mean = 5.69) is 1.23. The test reveals a statistical difference and H3b is therefore supported on a 1% significance level. The mean for the control group is between the low and high group and a statistical difference is found between the control group and the high group.

Table 37. Purchase intentions

WTB	Low		Control		High		Mean difference	Sig.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
H4b	3.80	1.52	-	-	5.40	1.47	1.60	.000***
	3.80	1.52	4.63	1.75	-	-	0.83	.017*
	-	-	4.63	1.75	5.40	1.47	0.77	.048*

* Significant at p< .05, **Significant at p< .01, *** Significant at p< .001.

For the variable purchase intentions, the mean difference between the low rating (mean = 3.80) and high rating (mean = 5.40) is 1.60. The test reveals a statistical difference and H4b is therefore supported on a 1% significance level. The mean for the control group is between the low and high group and a statistical difference is found between both the control group and the low group as well as the control group and the high group.

6.2. Summary of results - Study 2

Table 38. Support of hypotheses

	Hypothesis	
H1b	Receiving a high rating will generate higher *fulfillment of expectations* compared to receiving a low rating in a post experience setting.	Supported
H2b	Receiving a high rating will generate higher *rating* compared to receiving a low rating in a post experience setting.	Supported
H3b	Receiving a high rating will generate higher *satisfaction* compared to receiving a low rating in a post experience setting.	Supported
H4b	Receiving a high rating will generate higher *purchase intentions* compared to receiving a low rating in a post experience setting.	Supported

6.3. Concluding remarks - Study 2

Study 2 shows support for all of our b-hypotheses using a One-way Anova test, we have thus found support for that perceived consumer experience can indeed be influenced by ratings in

a post-experience setting, which was part of our research question. Furthermore, as our two studies were conducted in different contexts (tasting context and a film context), we have indications that our findings holds for different contexts as well.

7. Discussion

This chapter aims to discuss the results of our two studies and identify the implications of their results both theoretically and managerially, with the goal of justifying its importance while still being transparent about the limitations of the thesis. We end this chapter by suggesting where future research efforts could be directed to further explore the field of the influence of ratings.

7.1. Conclusions for Study 1 and 2

We previously defined fulfillment of expectations, rating, satisfaction and purchase intentions as constituents of "perceived consumer experiences", with basis in literature (e.g. Jones, Mothersbaugh & Betty, 2000; Mudambi & Schuff, 2010; Oliver, 1980), and used them as key indicators of change in the perceived consumer experience in both of our studies. The pre- and post-experience setting was achieved by letting the respondents receive a rating, i.e. the stimuli, either before or after going through an experience. We derived from literature a predicted influence of ratings, stating that consumers would evaluate our key indicators of change, and thus their perceived experience in the direction of the rating received. As predicted according to our theoretical framework, we found support for ratings' influence on perceived consumer experiences as hypothesized in **Figure 1** and **Figure 2**, included once more below. We have argued for that our findings are strong enough also in the pre-experience setting that was tested using non-parametric tests, thanks to our findings regarding the post-experience setting in two different contexts, i.e. the taste experience context as well as the film experience context that showed similar results. Our findings also go in line with consistency theories, by showing that respondents relieve inconsistency within their thoughts, i.e. between the rating received and what they would rate the experience, by rating it more in the direction of the rating received compared to the respondents receiving a rating in the opposite direction. We thus argue that we have answered to our research question in full: "Can perceived consumer experiences be influenced by ratings in both a pre- and post-experience setting?".

Figure 1. Supported influence of ratings on the variables defined as the perceived consumer experience in the pre-experience setting.
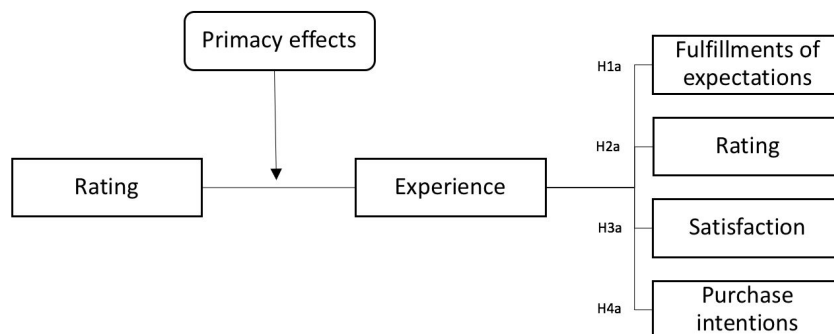
Figure 2. Supported influence of ratings on the variables defined as the perceived consumer experience in the post-experience setting.
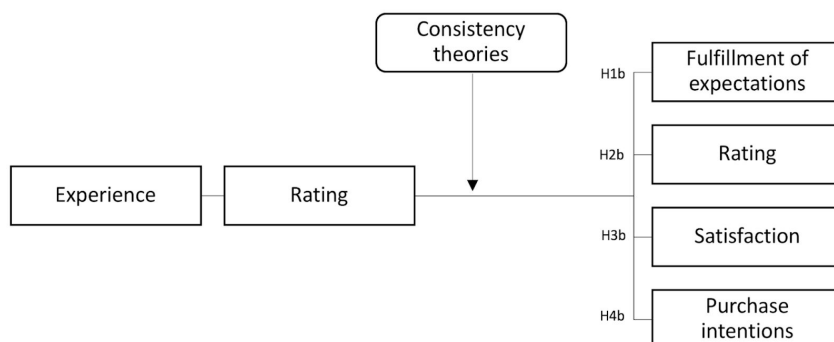
7.2. Implications

Theoretically and phenomenologically, this thesis is important as it emphasizes ratings' influence on perceived consumer experience rather than their impact on e.g. sales volumes. As companies use ratings in their communication about offerings as a way to attract customers, the managerial interest of this study should also be high. In the following subsections, we dig deeper into these implications, starting with the theoretical ones.

7.2.1 Theoretical implications

A lot of pre-existing research relating to reviews and ratings concerns purchase intentions and their impact on sales. A lot less put emphasis on their actual impact on perceived consumer

experiences, i.e. fulfillment of expectations, rating, and satisfaction, as we have defined it for the purpose of this thesis. Phenomenologically speaking, this thesis is therefore theoretically relevant, important and unique.

Furthermore, backward-framing as a phenomenon has been explored in the context of advertising information (Braun, 1999; Braun-LaTour, LaTour, Pickrell & Loftus, 2004; Braun-LaTour & La-Tour 2005; Braun-LaTour, Grinley & Loftus, 2006), but this thesis shed light on backward-framing also in the context of ratings. One may argue that ratings may indeed be used as advertising information. By definition they are not the same, especially since ratings can be less managerially controlled than other types of advertising information, but for the field of transformational advertising, the findings of this thesis are nonetheless important.

7.2.2. Managerial implications

Our study does not only provide implications to the theoretical landscape but the findings also give managers several highly important and valuable implications. Our research findings suggests how marketing managers can use ratings in their marketing communication to influence potential and current customers. Mentioned below are a few hands-on implications that managers, in many different business fields, can take action on starting today.

Firstly, ratings positively influences customers' perceived fulfillment of expectations, rating, satisfaction and purchase intentions. Low ratings will in contrast negatively influence the same variables. Exposure of ratings for products or services is not always something companies have full control over and our study thus shed light on the importance of making an effort to provide an offering that customers value and rate high, as this, will affect other customers' perceptions of the offering and their perceived experience of it in the long run and not just purchase intentions in the short run.

Secondly, even though high ratings have positive effects, managers should be careful. We have shown that customers who do not find the rating credible are not influenced, at least not in the direction of the rating, on the variables mentioned above. Therefore, an important finding is that managers opting for "too" high ratings, will risk reducing their potential impact on customers' perceived experience.

Third and lastly, ratings affect consumers' perceived experiences if exposed to them before the experience, but also after their experience. This means that managers can alter consumers' memory of an experience by exposing them to a rating post-purchase. Therefore, companies should not only focus on providing advertising information, like ratings, in a pre-purchase setting but also be sure to reach current customers with this type of informations after their gone-through customer experiences as they are still possible to influence.

7.3. Discussion of limitations

As with all research, there are limitations of both of our studies. For the pre-experience setting in the first experiment, i.e. the taste experiment, we found support for the hypothesized influence of ratings. However, this was done using non-parametric tests as the groups were too small for a One-way Anova test after they had been sorted according to the credibility confound check. The small sample sizes used to support our hypotheses in study 1 is a limitation to drawing generalizable conclusions. As previously argued however, we maintain that seeing the same result through this test, as with the One-way Anova test for the post-experience setting in the film experiment in Study 2, where both the product to be experienced, as well as the temporal setting and the degree to which the stimuli have been manipulated, have been altered, argues for a indication that such non-parametric test findings would hold also if the One-way Anova test could have been executed.

Furthermore, both the taste experience, i.e. the chocolate, and the film experience can be consumed for a relatively low price. Although the hedonic aspect of the experiences may be argued to be of different degree, which would support that our findings hold across different product categories, one may argue that they may not hold for more high-involvement purchases, e.g. cars. Previous research has as described in our theoretical framework stated that high-involvement consumers' opinions about a product were affected by negative reviews only if they perceived the quality of the review to be high (Lee, 2007). In this thesis we have argued for the choices we made to construct our stimuli, i.e. the manipulated ratings, and the only indication we have about the quality of these ratings is our credibility confound check. One may argue that this is the reason for why we did not see the hypothesized

influence of ratings on perceived consumer experiences when including the group that did not perceive the rating to be credible, in Study 1. As the sample was collected in the streets of Stockholm, there is reason to believe that only people that actually like chocolate would be interested enough to agree to participate in the experiment. These people may have thus have been too highly involved in the product category, i.e. chocolate, to perceive the rating to be credible, even if chocolate per se might not be argued to be an item subject to high-involvement purchases, compared to e.g. cars.

Also, some respondents did not consider the stimulus to be credible, especially in study 1, which resulted in these respondents being sorted out of the sample. There can be several reasons for this, a few worth mentioning could be that the ratings were not credible, asking the actual question made respondents think carefully about the credibility or simply chocolate bars are not that frequently rated in a similar fashion as movies, books, wine and concerts. Moreover, the facts that about 35% of the respondents had to be sorted out indicates a weakness for study 1.

7.4. Future Research

Although we argue to have found support for that our findings regarding the pre-experience setting, tested with non-parametric tests, would hold also if tested through a One-way Anova test, thanks to our findings regarding the post-experience setting that show a similar pattern of influence of ratings, we still encourage future research efforts to be directed towards strengthening these findings even further with larger samples and a One-way Anova test. Measuring the influence of ratings could potentially also be interesting.

The extent to which the sample of Study 1 was reduced due the confounding variable of credibility, motivates further research in the field of credibility of ratings and reviews. A lot of research has been done, especially focusing on what makes a review credible, i.e. the factors that increases the credibility of the review (e.g. Shan, 2016; Sparks, Perkins & Buckley, 2013; Guoqing, Kai & Fei, 2010), but what we would like to see, based on our findings, is further research on how people react to ratings that they do not perceive to be credible. In this thesis, we solely focus on the influence on perceived consumer experiences,

i.e. not on potential counterbalanced influence on the same, exclusively for the group that do no perceive the rating to be credible. For managers, given that such a counterbalanced influence would exist, the importance of only displaying ratings that are credible would grow heavily. Given that research would be done concerning the influence of ratings for a sample that does not perceive the rating to be credible, it would also be interesting to compare the influence between groups that perceive the rating to be credible vs not credible, in order to measure the extent of the influence. This would however assume that the group that does not perceive the rating to be credible is counter-influenced to some extent by the rating.

Our research question did not incorporate a dimension of *why* an influence of ratings would occur, neither a dimension motivating a comparison of the influence of ratings in the two temporal settings. However, already in our theoretical framework, we used theories about the primacy effect and the recency effect, as well as consistency theory, to further motivate our hypotheses regarding the pre- and post-experience setting, i.e. *why* we predicted the influence of ratings to occur in both of these two settings, but separately. As both a primacy effect and a recency effect have been shown in literature to be able to occur (Castel, 2008; Steiner & Rain, 1989), an experiment combining the pre- and post-experience setting for all respondents is therefore justified, i.e. letting all consumers receive a review both before and after the experience. If primed with one rating pre-experience, and given *another* rating post-experience, what influence would the ratings then have on perceived consumer experiences? Would the primacy effect and the recency effect cancel each other out, or would it just lead to the consumer not thinking any of the ratings are credible? Would one effect be stronger than the other?

As a rating is a quantitative measure, such a study would be theoretically motivated to determine which effect is greatest, but also in an authentic setting it is interesting as products, services, experiences, companies etcetera more often than not are judged through several rating processes, e.g. on different online sites, leading to a consumer being exposed to several, potentially conflicting, ratings and potentially in different temporal settings, pre- or post-experience.

Braun-LaTour, Grinley & Loftus (2006) also showed that repetition of false advertising information led to increased false memory creation in the post-experience setting, which motivates a study where the consumer receives *the same* rating both before and after the experience, in order to ascertain if that is the case also with multiple repetitions of false ratings in different timely settings, i.e. if false memory creation can be initiated already in the pre-experience setting, through priming the consumer, and augmented in the post-experience setting, to an extent larger than it would have been if the rating was received in the post-experience setting alone.

8. References

8.1. Academic sources

Amblee, N., & Bui, T. (2007). Freeware downloads: An empirical investigation into the impact of expert and user reviews on demand for digital goods. *AMCIS 2007 Proceedings*, 21.

Bell, E., & Thorpe, R. (2013). *A very short, fairly interesting and reasonably cheap book about management research*. Sage.

Braun, K. A. (1999). Postexperience advertising effects on consumer memory. *Journal of Consumer Research*, *25*(4), 319-334.

Braun-LaTour, K. A., Grinley, M. J., & Loftus, E. F. (2006). Tourist memory distortion. *Journal of Travel Research*, *44*(4), 360-367.

Braun-LaTour, K. A., & LaTour, M. S. (2005). Transforming consumer experience: When timing matters. *Journal of Advertising*, *34*(3), 19-30.

Braun-LaTour, K. A., LaTour, M. S., Pickrell, J. E. & Loftus, E. F. (2004). How and when advertising can influence memory for consumer experience. *Journal of Advertising*, *33*(4), 7-25.

Bryman, A., & Bell, E. (2007). Business research strategies. *Business research methods*.

Bryman, A., & Bell, E. (2011). *Företagsekonomiska forskningsmetoder*. Liber ekonomi.

Bryman, A. & Bell, E. 2015a, *Business research methods,* Oxford University Press, USA.

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, *36*(2), 429-437.

Cervellon, M. C., & I. Carey, L. (2014). Sustainable, hedonic and efficient: interaction effects between product properties and consumer reviews on post-experience responses. *European Journal of Marketing*, *48*(7/8), 1375-1394.

Chapman, K. J., & Aylesworth, A. (1999). Riding the coat-tails of a positive review: rave reviews and attitude transfer. *Journal of Consumer Marketing*, *16*(5), 418-440.

Cialdini, R. B. (1987). *Influence* (Vol. 3). Port Harcourt: A. Michel.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, *35*(3), 124-129.

Cromwell, H. (1950). The relative effect on audience attitude of the first versus the second argumentative speech of a series. *Communications Monographs*, *17*(2), 105-122.

Cronin Jr, J. J., & Taylor, S. A. (1992). Measuring service quality: a reexamination and extension. *The journal of marketing*, 55-68.

Fisher, R. A. (1992). The arrangement of field experiments. In *Breakthroughs in statistics* (pp. 82-91). Springer, New York, NY.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291-313.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-economics*, *40*(1), 35-42.

Guoqing, G., Kai, C., & Fei, H. (2010). An Empirical Study on the Influence of Perceived Credibility of Online Consumer Reviews [J]. *Contemporary Economy & Management*, *10*, 007.

Hall, B. F. (2002). A new model for measuring advertising effectiveness. *Journal of Advertising Research*, *42*(2), 23-31.

Hansen, L. F. (1925). The Psychology of Belief IV: The Law of Primacy in Persuasion.". *Journal of Abnormal Social Psychology*, *20*, 183-191.

Hao, Y., Ye, Q., Li, Y., & Cheng, Z. (2010, January). How does the valence of online consumer reviews matter in consumer decision making? Differences between search goods and experience goods. In *System sciences (HICSS), 2010 43rd hawaii international conference on* (pp. 1-10). IEEE.

Hilger, J., Rafert, G., & Villas-Boas, S. (2011). Expert opinion and the demand for experience goods: an experimental approach in the retail wine market. *Review of Economics and Statistics*, *93*(4), 1289-1296.

Hoch, S. J., & Deighton, J. (1989). Managing what consumers learn from experience. *The Journal of Marketing*, 1-20.

Hoch, S. J., & Ha, Y. W. (1986). Consumer learning: Advertising and the ambiguity of product experience. *Journal of consumer research*, *13*(2), 221-233.

Hoch, S. J. (2002). Product experience is seductive. *Journal of Consumer Research*, *29*(3), 448-454.

Holbrook, M. B., & Batra, R. (1987). Assessing the role of emotions as mediators of consumer responses to advertising. *Journal of consumer research*, *14*(3), 404-420.

Jones, M. A., Mothersbaugh, D. L., & Beatty, S. E. (2000). Switching barriers and repurchase intentions in services. *Journal of retailing*, *76*(2), 259-274.

Kim, W. G., Ng, C. Y. N., & Kim, Y. S. (2009). Influence of institutional DINESERV on customer satisfaction, return intention, and word-of-mouth. *International Journal of Hospitality Management*, *28*(1), 10-17.

Lee, J., Park, D. H., & Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic commerce research and applications*, *7*(3), 341-352.

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of consumer research*, *15*(3), 374-378.

Liu, S. Q., Ozanne, M., & Mattila, A. S. (2018). Does expressing subjectivity in online reviews enhance persuasion?. *Journal of Consumer Marketing*, (just-accepted), 00-00.

Malhotra, N. K., & Birks, D. F. (2007). Marketing research: An applied approach: Pearson Education.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly*, 185-200.

MUNDRY, R., & FISCHER, J. (1998). Use of statistical programs for nonparametric tests of small samples often leads to incorrect Pvalues: examples from Animal Behaviour. *Animal behaviour*, *56*(1), 256-259.

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

Qiu, L., Pang, J., & Lim, K. H. (2012). Effects of conflicting aggregated rating on eWOM review credibility and diagnosticity: The moderating role of review valence. *Decision Support Systems*, *54*(1), 631-643.

Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour, 11*(3), 234-243.

Shan, Y. (2016). How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation. *Computers in Human Behavior*, *55*, 633-641.

Sparks, B. A., Perkins, H. E., & Buckley, R. (2013). Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, *39*, 1-9.

Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology*, *38*(6), 841.

Steiner, D. D., & Rain, J. S. (1989). Immediate and delayed primacy and recency effects in performance evaluation. *Journal of Applied Psychology*, *74*(1), 136.

Söderlund, M. (2005). *Mätningar och mått: i marknadsundersökarens värld*. Liber ekonomi.

Söderlund, M. (2010). *Experiment med människor*. Liber.

Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of marketing*, *73*(5), 90-102.

Vakratsas, D., & Ambler, T. (1999). How advertising works: what do we really know?. *The Journal of Marketing*, 26-43.

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, *30*(1), 123-127.

Wu, L., Shen, H., Li, M., & Deng, Q. (2017). Sharing information now vs later: The effect of temporal contiguity cue and power on consumer response toward online reviews. *International Journal of Contemporary Hospitality Management*, *29*(2), 648-668.

Wyatt, R. O., & Badger, D. P. (1984). How reviews affect interest in and evaluation of films. *Journalism Quarterly*, *61*(4), 874-878.

Yang, J., Sarathy, R., & Walsh, S. M. (2016). Do review valence and review volume impact consumers' purchase decisions as assumed?. *Nankai Business Review International*, *7*(2), 231-257.

Zhu, F., & Zhang, X. (2006). The influence of online consumer reviews on the demand for experience goods: The case of video games. *ICIS 2006 Proceedings*, 25.

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.

8.2. Non-academic sources

O'Connor, K & Cheema, A. (2018, October 3). Why Ratings on Everything from Wine to Amazon Products Improve Over Time. Retrieved December, 10, 2018 from https://hbr.org/2018/10/research-why-ratings-on-everything-from-wine-to-amazon-products-improve-over-time

9. Appendix

9.1. Questionnaire - Pre-test

Betygsätt Plopp Gott & Blandat (1=Sämst, 10=Bäst)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Har du smakat Plopp Gott & Blandat tidigare?

Ja

Nej

Betygsätt Toms choklad med Ferrari (1=Sämst, 10=Bäst)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Har du smakat Toms choklad med Ferrari tidigare?

Ja

Nej

Betyg: **2,4** / 10

⭐⭐🌟☆☆☆☆☆☆☆

<small>(Baserat på 3567 omdömen)</small>

Betyg: **9,6** / 10

⭐⭐⭐⭐⭐⭐⭐⭐⭐🌟

<small>(Baserat på 3567 omdömen)</small>

Hej,

Vi är två studenter från Handelshögskolan i Stockholm. Denna enkätundersökning som du deltar i ligger till grund för vårt examensarbete. I enkäten finns 19 frågor, att medverka tar cirka 3-5 minuter och ditt deltagande är anonymt. Undersökningen innehåller ett smakprov av en chokladkaka, vänligen meddela oss om eventuella allergier.

I enkäten finns vissa instruktioner, exempelvis om när du ska smaka chokladkakan, vänligen följ dessa noga.

Tack för din tid och ditt engagemang!
Eleonor & Erik

En sak bara innan vi börjar…

Vad är dina förväntningar på chokladkakan innan du smakar på den?

	1	2	3	4	5	6	7	
Negativa	○	○	○	○	○	○	○	Positiva
Väldigt låga	○	○	○	○	○	○	○	Väldigt höga
Ofördelaktiga	○	○	○	○	○	○	○	Fördelaktiga

Nu kan du få smaka på chokladkakan!

☺

2. Hur var smakupplevelsen jämfört med dina förväntningar?

	1	2	3	4	5	6	7	
Mycket sämre än förväntat	○	○	○	○	○	○	○	Mycket bättre än förväntat

3. Vänligen betygsätt chokladkakan (1=Sämst, 10=Bäst)

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

4. Jag känner mig självsäker på betyget jag satte

	1	2	3	4	5	6	7	
Instämmer inte alls	○	○	○	○	○	○	○	Instämmer helt och hållet

5. Hur tror du att du betygsatte chokladkakan i jämförelse med andra?

Mycket lägre			Ungefär lika			Mycket högre
○	○	○	○	○	○	○

6. Utvärdera din smakupplevelse

	1	2	3	4	5	6	7	
Väldigt missnöjd	○	○	○	○	○	○	○	Väldigt nöjd
Mindre lyckad	○	○	○	○	○	○	○	Lyckad
Väldigt otillfredsställande	○	○	○	○	○	○	○	Väldigt tillfredsställande

7. Utvärdera kvaliteten på chokladkakan

	1	2	3	4	5	6	7	
Låg kvalitet	○	○	○	○	○	○	○	Hög kvalitet
Sämre än genomsnittet	○	○	○	○	○	○	○	Bättre än genomsnittet
Sämre än konkurrenterna	○	○	○	○	○	○	○	Bättre än konkurrenterna

8. I vilken utsträckning smakade chokladkakan "Gott och Blandat"?

	1	2	3	4	5	6	7	
Väldigt lite	○	○	○	○	○	○	○	Väldigt mycket

9. Jag är villig att berätta positiva saker om chokladkakan för andra

	1	2	3	4	5	6	7	
Instämmer inte alls	○	○	○	○	○	○	○	Instämmer helt och hållet

10. Om du skulle köpa snacks, vad är sannolikheten att du skulle överväga att köpa denna chokladkakan?

	1	2	3	4	5	6	7	
Otroligt	○	○	○	○	○	○	○	Troligt
Osannolikt	○	○	○	○	○	○	○	Sannolikt
Omöjligt	○	○	○	○	○	○	○	Möjligt

11. Hur mycket skulle du vara villig att betala för den här chokladkakan? (75g, skriv i kronor)

_____

12. Estimera temperaturen i rummet du befinner dig just nu (Skriv i grader Celsius)

_____

13. Du såg en bild på chokladkakan i början av enkäten, vilket betyg hade denna chokladkakan fått på bilden?

○ 2,4 / 10

○ 9,6 / 10

○ Jag vet inte

14. Betyget som jag valde i föregående fråga är generellt ett...

○ Lågt betyg

○ Högt betyg

○ Jag vet inte

15. Var betyget trovärdigt?

○ Ja

○ Nej

16. Har du smakat denna chokladkaka tidigare?

○ Ja

○ Nej

17. Din ålder (Skriv i antal år, exempelvis 35)

_____

18. Din könstillhörighet

○ Kvinna

○ Man

○ Annan

○ Jag vill inte svara

Betyg: **5,4** / 10

★★★★★☆☆☆☆☆

(Baserat på 3567 omdömen)

Betyg: **8,6** / 10

★★★★★★★★☆☆

(Baserat på 3567 omdömen)

9.5. Survey flow - Study 2
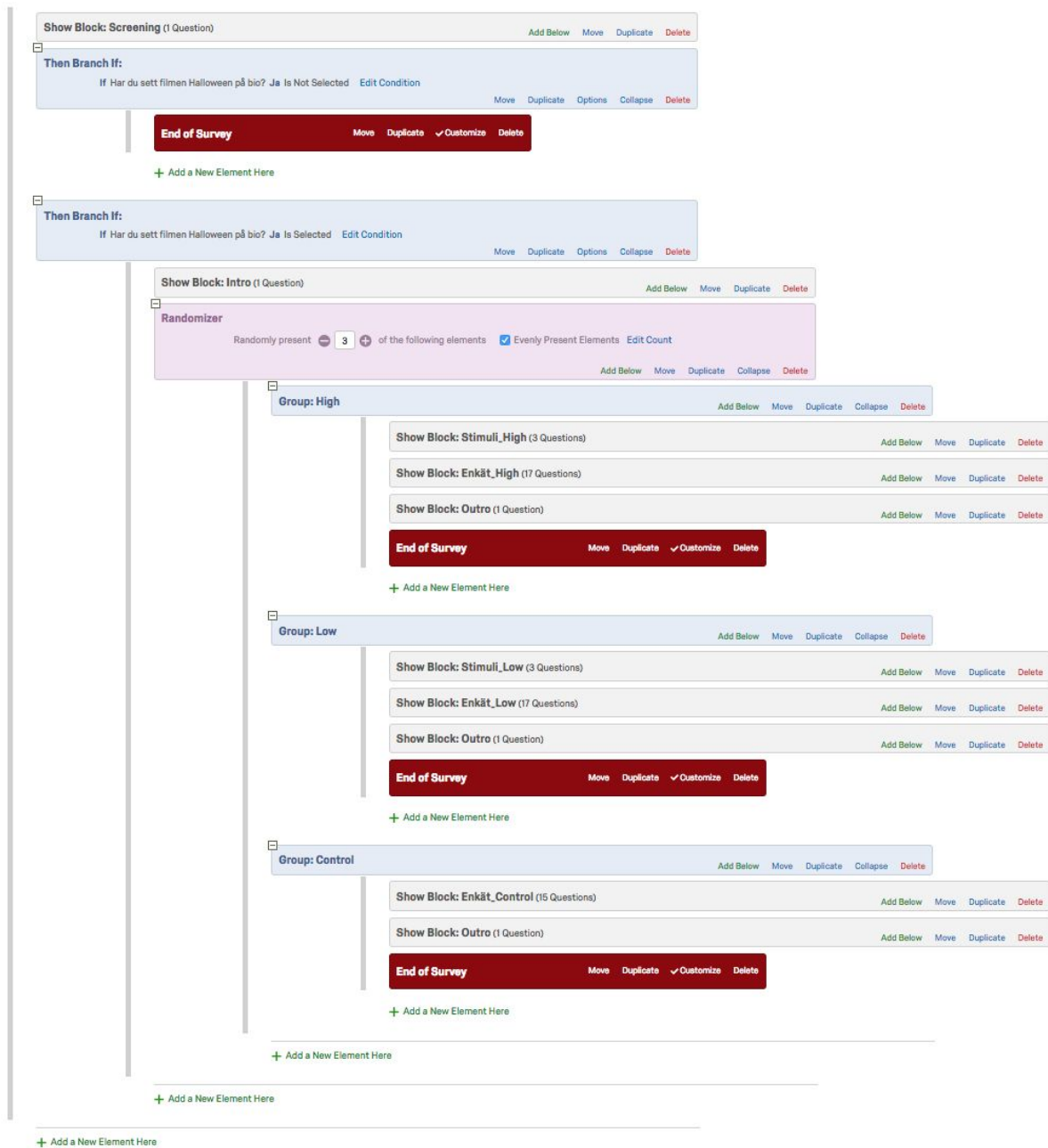
9.6. Questionnaire - Study 2

Hej,

Du var nyligen och såg filmen Halloween på bio och vi skulle gärna vilja veta dina tankar om filmen och ditt biobesök.

Att fylla i enkäten tar cirka 3-5 minuter och ditt deltagande är anonymt.

Undersökningen är en del av ett experiment som görs i forskningssyfte.

Tack för din tid!

Vad var dina förväntningar på filmen innan du såg den?

Negativa	O O O O O O	Positiva
Väldigt låga	O O O O O O	Väldigt höga
Ofördelaktiga	O O O O O O	Fördelaktiga

Hur var filmupplevelsen jämfört med dina förväntningar?

Mycket sämre än förväntat   O O O O O O   Mycket bättre än förväntat

Vänligen betygsätt filmen (1=Sämst, 10=Bäst)

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Jag känner mig självsäker på betyget jag satte

Instämmer inte alls   O O O O O O   Instämmer helt och hållet

Hur tror du att du betygsatte filmen i jämförelse med andra?

Mycket lägre   O O O O O O   Mycket Högre

Utvärdera din filmupplevelse

Väldigt missnöjd   O O O O O O   Väldigt nöjd

Mindre lyckad   O O O O O O   Lyckad

Väldigt otillfredsställande   O O O O O O   Väldigt tillfredsställande

Utvärdera kvaliteten på filmen

Låg kvalitet   O O O O O O   Hög kvalitet

Sämre än genomsnittet   O O O O O O   Bättre än genomsnittet

Sämre än filmer i samma genre   O O O O O O   Bättre än filmer i samma genre

Jag upplevde att filmen var...

Inte alls läskig   O O O O O O   Väldigt läskig

Jag är villig att berätta positiva saker om filmen för andra

Instämmer inte alls ○ ○ ○ ○ ○ ○ Instämmer helt och hållet

Hur skulle du känna inför en uppföljare till filmen?

Vill inte se ○ ○ ○ ○ ○ ○ Vill se

Likgiltig ○ ○ ○ ○ ○ ○ Nyfiken

Ointresserad ○ ○ ○ ○ ○ ○ Intresserad

Uppskatta temperaturen i rummet du befinner dig i just nu (skriv i grader Celsius)

I början av enkäten såg du en bild på filmen med ett betyg, vilket betyg hade filmen fått?

5,4 / 10

8,6 / 10

Jag vet inte

Betyget som jag valde i föregående fråga är för en film generellt ett...

Lågt betyg

Högt betyg

Jag vet inte

Var betyget på bilden trovärdigt?

Ja

Nej

Vilken film var det du såg?

Halloween

A simple favor

Jag vet inte

Har du sett andra recensioner/betyg på just denna film?

Ja

Nej

Din ålder (skriv i antal år, exempelvis 35)

Din könstillhörighet

Kvinna

Man

Annan

Jag vill inte svara

Tack för att du var med i vår undersökning!

Det är viktigt för oss att vara transparenta mot dig som biobesökare. Därför vill vi vara tydliga med att detta var ett experiment och att betyget du fick se i början av undersökningen inte var ett riktigt betyg utan ett manipulerat betyg. Detta var i forskningssyfte för att se hur biobesökare reagerar på olika slags betyg. Om du vill se ett verkligt betyg på filmen rekommenderar vi dig att gå in på exempelvis www.IMDB.com.

Klicka på pilen nedan för att avsluta enkäten!

9.7. Raw data - Pre-test

Statistics

		Betygsätt Plopp Gott & Blandat (1=Sämst, 10=Bäst)	Har du smakat Plopp Gott & Blandat tidigare? (1=Ja 2= Nej)	Betygsätt Toms choklad med Ferrari (1=Sämst, 10=Bäst)	Har du smakat Toms choklad med Ferrari tidigare? (1=Ja 2=Nej)
N	Valid	35	35	35	35
	Missing	0	0	0	0
Mean		6.14	2.00	5.34	2.00
Std. Deviation		1.734	.000	1.781	.000
Minimum		2	2	2	2
Maximum		10	2	8	2

Frequency Table

Betygsätt Plopp Gott & Blandat (1=Sämst, 10=Bäst)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	1	2.9	2.9	2.9
	3	1	2.9	2.9	5.7
	4	3	8.6	8.6	14.3
	5	7	20.0	20.0	34.3
	6	8	22.9	22.9	57.1
	7	10	28.6	28.6	85.7
	8	2	5.7	5.7	91.4
	9	1	2.9	2.9	94.3
	10	2	5.7	5.7	100.0
	Total	35	100.0	100.0	

Har du smakat Plopp Gott & Blandat tidigare? (1=Ja 2= Nej)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nej	35	100.0	100.0	100.0

Betygsätt Toms choklad med Ferrari (1=Sämst, 10=Bäst)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	4	11.4	11.4	11.4
	3	1	2.9	2.9	14.3
	4	5	14.3	14.3	28.6
	5	8	22.9	22.9	51.4
	6	7	20.0	20.0	71.4
	7	6	17.1	17.1	88.6
	8	4	11.4	11.4	100.0
	Total	35	100.0	100.0	

Har du smakat Toms choklad med Ferrari tidigare? (1=Ja 2=Nej)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Nej	35	100.0	100.0	100.0

9.8. Raw data - Study 1

9.8.1. Demography and group assignment

Statistics

		What was the manipulation?	How old are you?	What is your gender?
N	Valid	167	167	167
	Missing	0	0	0
Mean		2.92	33.02	1.51
Std. Deviation		1.452	12.106	.513
Minimum		1	18	1
Maximum		5	74	3

Frequency Table

What was the manipulation?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Controll Group	40	24.0	24.0	24.0
	Before Low	30	18.0	18.0	41.9
	Before High	32	19.2	19.2	61.1
	After Low	33	19.8	19.8	80.8
	After High	32	19.2	19.2	100.0
	Total	167	100.0	100.0	

What is your gender?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	83	49.7	49.7	49.7
	Male	83	49.7	49.7	99.4
	Other	1	.6	.6	100.0
	Total	167	100.0	100.0	

9.8.2. Cronbach's Alpha

Satisfaction

Reliability Statistics

Cronbach's Alpha	N of Items
.958	3

Item Statistics

	Mean	Std. Deviation	N
Evaluate your taste experience	4.78	1.354	167
Evaluate your taste experience	4.76	1.478	167
Evaluate your taste experience	4.65	1.405	167

Purchase Intentions

Reliability Statistics

Cronbach's Alpha	N of Items
.960	3

Item Statistics

	Mean	Std. Deviation	N
If you where buying snacks, what is the probability you would buy this chocolate?	3.19	1.842	167
If you where buying snacks, what is the probability you would buy this chocolate?	3.28	1.891	167
If you where buying snacks, what is the probability you would buy this chocolate?	3.93	1.877	167

9.8.3. One-way ANOVA

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
How was the experience according to your expectations?	Controll Group	40	4.13	1.381	.218	3.68	4.57	1	7
	Before Low	30	4.80	1.349	.246	4.30	5.30	2	7
	Before High	32	5.22	1.385	.245	4.72	5.72	2	7
	After Low	33	4.61	1.391	.242	4.11	5.10	2	7
	After High	32	5.34	1.208	.214	4.91	5.78	3	7
	Total	167	4.78	1.406	.109	4.57	5.00	1	7
Rate the chocolate	Controll Group	40	5.68	1.859	.294	5.08	6.27	1	8
	Before Low	30	6.03	2.236	.408	5.20	6.87	2	10
	Before High	32	6.59	2.061	.364	5.85	7.34	2	9
	After Low	33	5.76	1.786	.311	5.12	6.39	2	9
	After High	32	6.75	1.934	.342	6.05	7.45	2	9
	Total	167	6.14	1.997	.155	5.83	6.44	1	10
Evaluate your taste experience	Controll Group	40	4.3833	1.25530	.19848	3.9819	4.7848	1.00	6.67
	Before Low	30	4.7556	1.46983	.26835	4.2067	5.3044	2.00	7.00
	Before High	32	5.0000	1.40148	.24775	4.4947	5.5053	1.67	7.00
	After Low	33	4.4141	1.35408	.23572	3.9340	4.8943	1.67	7.00
	After High	32	5.2083	1.19362	.21100	4.7780	5.6387	2.00	6.67
	Total	167	4.7325	1.35684	.10500	4.5252	4.9398	1.00	7.00
If you where buying snacks, what is the probability you would buy this chocolate?	Controll Group	40	3.2333	1.61933	.25604	2.7154	3.7512	1.00	6.67
	Before Low	30	3.4333	1.87778	.34283	2.7322	4.1345	1.00	7.00
	Before High	32	3.8438	1.84135	.32551	3.1799	4.5076	1.00	6.67
	After Low	33	3.0606	1.76079	.30651	2.4363	3.6850	1.00	7.00
	After High	32	3.8229	1.90262	.33634	3.1369	4.5089	1.00	7.00
	Total	167	3.4651	1.79914	.13922	3.1902	3.7399	1.00	7.00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
How was the experience according to your expectations?	Between Groups	34.498	4	8.625	4.756	.001
	Within Groups	293.741	162	1.813		
	Total	328.240	166			
Rate the chocolate	Between Groups	32.311	4	8.078	2.079	.086
	Within Groups	629.521	162	3.886		
	Total	661.832	166			
Evaluate your taste experience	Between Groups	17.772	4	4.443	2.501	.045
	Within Groups	287.836	162	1.777		
	Total	305.609	166			
If you where buying snacks, what is the probability you would buy this chocolate?	Between Groups	16.263	4	4.066	1.264	.286
	Within Groups	521.061	162	3.216		
	Total	537.324	166			

Scheffe

Dependent Variable	(I) What was the manipulation?	(J) What was the manipulation?	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
How was the experience according to your expectations?	Controll Group	Before Low	−.675	.325	.370	−1.69	.34
		Before High	−1.094*	.319	.022	−2.09	−.10
		After Low	−.481	.317	.680	−1.47	.51
		After High	−1.219*	.319	.007	−2.21	−.22
	Before Low	Controll Group	.675	.325	.370	−.34	1.69
		Before High	−.419	.342	.827	−1.49	.65
		After Low	.194	.340	.988	−.86	1.25
		After High	−.544	.342	.641	−1.61	.52
	Before High	Controll Group	1.094*	.319	.022	.10	2.09
		Before Low	.419	.342	.827	−.65	1.49
		After Low	.613	.334	.501	−.43	1.65
		After High	−.125	.337	.998	−1.17	.92
	After Low	Controll Group	.481	.317	.680	−.51	1.47
		Before Low	−.194	.340	.988	−1.25	.86
		Before High	−.613	.334	.501	−1.65	.43
		After High	−.738	.334	.305	−1.78	.30
	After High	Controll Group	1.219*	.319	.007	.22	2.21
		Before Low	.544	.342	.641	−.52	1.61
		Before High	.125	.337	.998	−.92	1.17
		After Low	.738	.334	.305	−.30	1.78
Rate the chocolate	Controll Group	Before Low	−.358	.476	.966	−1.84	1.13
		Before High	−.919	.468	.428	−2.38	.54
		After Low	−.083	.464	1.000	−1.53	1.36
		After High	−1.075	.468	.264	−2.53	.38
	Before Low	Controll Group	.358	.476	.966	−1.13	1.84
		Before High	−.560	.501	.869	−2.12	1.00
		After Low	.276	.497	.989	−1.27	1.83
		After High	−.717	.501	.727	−2.28	.84
	Before High	Controll Group	.919	.468	.428	−.54	2.38
		Before Low	.560	.501	.869	−1.00	2.12
		After Low	.836	.489	.572	−.69	2.36
		After High	−.156	.493	.999	−1.69	1.38
	After Low	Controll Group	.083	.464	1.000	−1.36	1.53
		Before Low	−.276	.497	.989	−1.83	1.27
		Before High	−.836	.489	.572	−2.36	.69
		After High	−.992	.489	.394	−2.52	.53
	After High	Controll Group	1.075	.468	.264	−.38	2.53
		Before Low	.717	.501	.727	−.84	2.28
		Before High	.156	.493	.999	−1.38	1.69
		After Low	.992	.489	.394	−.53	2.52
Evaluate your taste experience	Controll Group	Before Low	−.37222	.32194	.855	−1.3754	.6310
		Before High	−.61667	.31614	.436	−1.6018	.3684
		After Low	−.03081	.31347	1.000	−1.0076	.9460
		After High	−.82500	.31614	.152	−1.8101	.1601
	Before Low	Controll Group	.37222	.32194	.855	−.6310	1.3754
		Before High	−.24444	.33875	.971	−1.3000	.8111
		After Low	.34141	.33625	.905	−.7064	1.3892
		After High	−.45278	.33875	.775	−1.5083	.6028
	Before High	Controll Group	.61667	.31614	.436	−.3684	1.6018
		Before Low	.24444	.33875	.971	−.8111	1.3000
		After Low	.58586	.33070	.537	−.4446	1.6164
		After High	−.20833	.33324	.983	−1.2467	.8301
	After Low	Controll Group	.03081	.31347	1.000	−.9460	1.0076
		Before Low	−.34141	.33625	.905	−1.3892	.7064
		Before High	−.58586	.33070	.537	−1.6164	.4446
		After High	−.79419	.33070	.223	−1.8247	.2363
	After High	Controll Group	.82500	.31614	.152	−.1601	1.8101
		Before Low	.45278	.33875	.775	−.6028	1.5083
		Before High	.20833	.33324	.983	−.8301	1.2467
		After Low	.79419	.33070	.223	−.2363	1.8247

94

If you where buying snacks, what is the probability you would buy this chocolate?	Controll Group	Before Low	−.20000	.43316	.995	−1.5497	1.1497
		Before High	−.61042	.42535	.725	−1.9358	.7150
		After Low	.17273	.42176	.997	−1.1415	1.4869
		After High	−.58958	.42535	.750	−1.9150	.7358
	Before Low	Controll Group	.20000	.43316	.995	−1.1497	1.5497
		Before High	−.41042	.45577	.937	−1.8306	1.0098
		After Low	.37273	.45242	.954	−1.0370	1.7825
		After High	−.38958	.45577	.947	−1.8098	1.0306
	Before High	Controll Group	.61042	.42535	.725	−.7150	1.9358
		Before Low	.41042	.45577	.937	−1.0098	1.8306
		After Low	.78314	.44495	.543	−.6033	2.1696
		After High	.02083	.44836	1.000	−1.3763	1.4180
	After Low	Controll Group	−.17273	.42176	.997	−1.4869	1.1415
		Before Low	−.37273	.45242	.954	−1.7825	1.0370
		Before High	−.78314	.44495	.543	−2.1696	.6033
		After High	−.76231	.44495	.570	−2.1488	.6242
	After High	Controll Group	.58958	.42535	.750	−.7358	1.9150
		Before Low	.38958	.45577	.947	−1.0306	1.8098
		Before High	−.02083	.44836	1.000	−1.4180	1.3763
		After Low	.76231	.44495	.570	−.6242	2.1488

9.8.4. Group assignment (only credible)

What was the manipulation?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Controll Group	40	37.0	37.0	37.0
	Before Low	13	12.0	12.0	49.1
	Before High	17	15.7	15.7	64.8
	After Low	22	20.4	20.4	85.2
	After High	16	14.8	14.8	100.0
	Total	108	100.0	100.0	

9.8.5. Non-parametric K independent sample - Kruskal-Wallis (only credible)

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of How was the experience according to your expectations? is the same across categories of What was the manipulation?.	Independent–Samples Kruskal–Wallis Test	.000	Reject the null hypothesis.
2	The distribution of Rate the chocolate is the same across categories of What was the manipulation?.	Independent–Samples Kruskal–Wallis Test	.000	Reject the null hypothesis.
3	The distribution of Evaluate your taste experience is the same across categories of What was the manipulation?.	Independent–Samples Kruskal–Wallis Test	.000	Reject the null hypothesis.
4	The distribution of If you where buying snacks, what is the probability you would buy this chocolate? is the same across categories of What was the manipulation?.	Independent–Samples Kruskal–Wallis Test	.000	Reject the null hypothesis.

Fulfillment of expectations

Pairwise Comparisons of What was the manipulation?
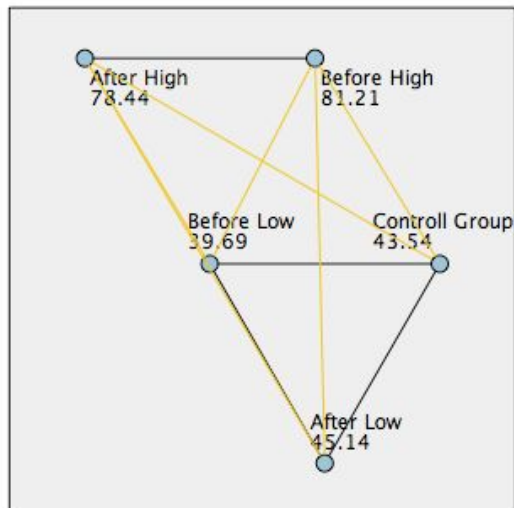
Each node shows the sample average rank of What was the manipulation?.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Before Low-Controll Group	3.845	9.800	.392	.695	1.000
Before Low-After Low	-5.444	10.738	-.507	.612	1.000
Before Low-After High	-38.745	11.462	-3.380	.001	.007
Before Low-Before High	-41.514	11.310	-3.671	.000	.002
Controll Group-After Low	-1.599	8.148	-.196	.844	1.000
Controll Group-After High	-34.900	9.080	-3.844	.000	.001
Controll Group-Before High	-37.668	8.887	-4.239	.000	.000
After Low-After High	-33.301	10.086	-3.302	.001	.010
After Low-Before High	36.070	9.912	3.639	.000	.003
After High-Before High	2.768	10.692	.259	.796	1.000

Rating

Pairwise Comparisons of What was the manipulation?

Each node shows the sample average rank of What was the manipulation?.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Before Low-After Low	-2.598	10.831	-.240	.810	1.000
Before Low-Controll Group	14.262	9.885	1.443	.149	1.000
Before Low-After High	-45.587	11.561	-3.943	.000	.001
Before Low-Before High	-46.991	11.408	-4.119	.000	.000
After Low-Controll Group	11.664	8.218	1.419	.156	1.000
After Low-After High	-42.989	10.173	-4.226	.000	.000
After Low-Before High	44.393	9.998	4.440	.000	.000
Controll Group-After High	-31.325	9.159	-3.420	.001	.006
Controll Group-Before High	-32.729	8.964	-3.651	.000	.003
After High-Before High	1.404	10.785	.130	.896	1.000

Satisfaction

Pairwise Comparisons of What was the manipulation?
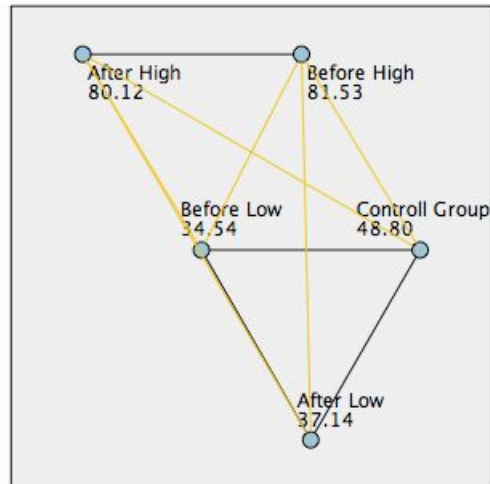

Each node shows the sample average rank of What was the manipulation?.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Before Low-After Low	-3.848	10.923	-.352	.725	1.000
Before Low-Controll Group	13.274	9.968	1.332	.183	1.000
Before Low-Before High	-43.814	11.504	-3.809	.000	.001
Before Low-After High	-46.337	11.659	-3.974	.000	.001
After Low-Controll Group	9.426	8.288	1.137	.255	1.000
After Low-Before High	39.967	10.083	3.964	.000	.001
After Low-After High	-42.489	10.259	-4.142	.000	.000
Controll Group-Before High	-30.540	9.040	-3.378	.001	.007
Controll Group-After High	-33.062	9.236	-3.580	.000	.003
Before High-After High	-2.522	10.876	-.232	.817	1.000

Purchase intentions
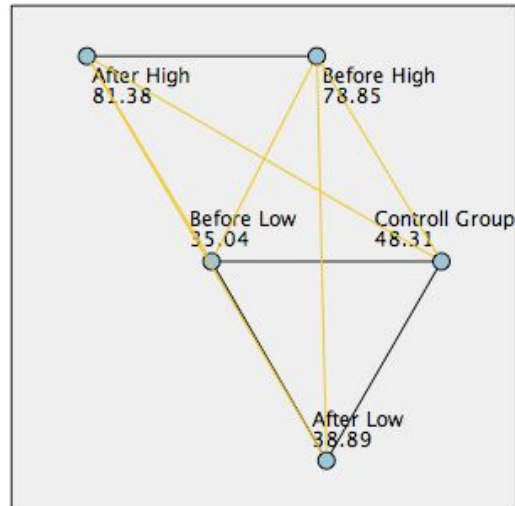
Pairwise Comparisons of What was the manipulation?

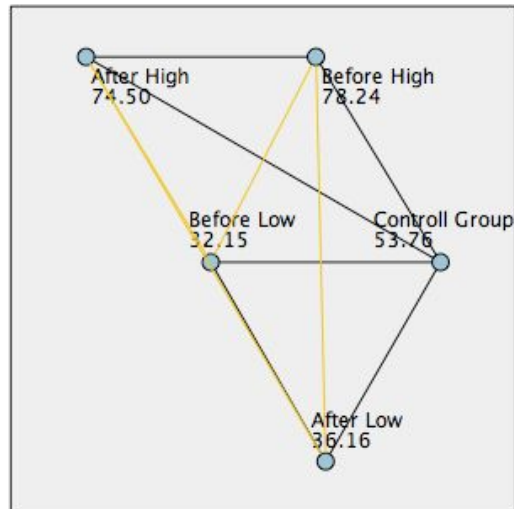
Each node shows the sample average rank of What was the manipulation?.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Before Low–After Low	−4.005	10.916	−.367	.714	1.000
Before Low–Controll Group	21.609	9.962	2.169	.030	.301
Before Low–After High	−42.346	11.651	−3.634	.000	.003
Before Low–Before High	−46.081	11.497	−4.008	.000	.001
After Low–Controll Group	17.603	8.282	2.125	.034	.336
After Low–After High	−38.341	10.252	−3.740	.000	.002
After Low–Before High	42.076	10.076	4.176	.000	.000
Controll Group–After High	−20.738	9.230	−2.247	.025	.247
Controll Group–Before High	−24.473	9.034	−2.709	.007	.068
After High–Before High	3.735	10.869	.344	.731	1.000

9.9. Raw data - Study 2

9.9.1. Demography and group assignment

Statistics

		Malipulation	Din ålder (skriv i antal år, exempelvis 35)	Din könstillhörigh et
N	Valid	170	170	170
	Missing	0	0	0
Mean		2.08	35.60	1.56
Std. Deviation		.795	10.750	.498
Minimum		1	18	1
Maximum		3	64	2

Frequency Table

Malipulation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	High	47	27.6	27.6	27.6
	Low	62	36.5	36.5	64.1
	Control	61	35.9	35.9	100.0
	Total	170	100.0	100.0	

Din könstillhörighet

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Kvinna	75	44.1	44.1	44.1
	Man	95	55.9	55.9	100.0
	Total	170	100.0	100.0	

9.9.2. Cronbach's Alpha

Satisfaction

Reliability Statistics

Cronbach's Alpha	N of Items
.923	3

Item Statistics

	Mean	Std. Deviation	N
Utvärdera din filmupplevelse – Väldigt missnöjd:Väldigt nöjd	5.04	1.249	170
Utvärdera din filmupplevelse – Mindre lyckad:Lyckad	5.00	1.332	170
Utvärdera din filmupplevelse – Väldigt otillfredsställande: Väldigt tillfredsställande	4.84	1.322	170

Purchase intentions

Reliability Statistics

Cronbach's Alpha	N of Items
.959	3

Item Statistics

	Mean	Std. Deviation	N
Hur skulle du känna inför en uppföljare till filmen? – Vill inte se:Vill se	4.49	1.824	170
Hur skulle du känna inför en uppföljare till filmen? – Likgiltig: Nyfiken	4.63	1.743	170
Hur skulle du känna inför en uppföljare till filmen? – Ointresserad: Intresserad	4.51	1.765	170

9.9.3. One-way ANOVA

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Hur var filmupplevelsen jämfört med dina förväntningar? – Mycket sämre än förväntat: Mycket bättre än förväntat	High	47	5.34	1.027	.150	5.04	5.64	3	7
	Low	62	4.27	1.230	.156	3.96	4.59	1	7
	Control	61	4.77	1.203	.154	4.46	5.08	1	7
	Total	170	4.75	1.236	.095	4.56	4.93	1	7
Vänligen betygsätt filmen (1=Sämst, 10=Bäst)	High	47	7.36	1.358	.198	6.96	7.76	3	10
	Low	62	5.47	1.647	.209	5.05	5.89	1	8
	Control	61	6.41	1.687	.216	5.98	6.84	1	10
	Total	170	6.33	1.750	.134	6.06	6.59	1	10
Experience_Index	High	47	5.6879	.98884	.14424	5.3976	5.9783	3.00	7.00
	Low	62	4.4624	1.21219	.15395	4.1545	4.7702	1.33	7.00
	Control	61	4.8962	1.10309	.14124	4.6137	5.1787	2.00	7.00
	Total	170	4.9569	1.21170	.09293	4.7734	5.1403	1.33	7.00
RePurchase_Index	High	47	5.4043	1.47103	.21457	4.9723	5.8362	1.00	7.00
	Low	62	3.8011	1.52270	.19338	3.4144	4.1878	1.00	7.00
	Control	61	4.6339	1.74664	.22363	4.1865	5.0812	1.00	7.00
	Total	170	4.5431	1.70886	.13106	4.2844	4.8019	1.00	7.00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Hur var filmupplevelsen jämfört med dina förväntningar? – Mycket sämre än förväntat: Mycket bättre än förväntat	Between Groups	30.445	2	15.222	11.165	.000
	Within Groups	227.679	167	1.363		
	Total	258.124	169			
Vänligen betygsätt filmen (1=Sämst, 10=Bäst)	Between Groups	96.512	2	48.256	19.140	.000
	Within Groups	421.041	167	2.521		
	Total	517.553	169			
Experience_Index	Between Groups	40.506	2	20.253	16.290	.000
	Within Groups	207.622	167	1.243		
	Total	248.128	169			
RePurchase_Index	Between Groups	69.495	2	34.747	13.685	.000
	Within Groups	424.022	167	2.539		
	Total	493.517	169			

Multiple Comparisons

Scheffe

Dependent Variable	(I) Malipulation	(J) Malipulation	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Hur var filmupplevelsen jämfört med dina förväntningar? – Mycket sämre än förväntat: Mycket bättre än förväntat	High	Low	1.066*	.226	.000	.51	1.62
		Control	.570*	.227	.045	.01	1.13
	Low	High	−1.066*	.226	.000	−1.62	−.51
		Control	−.496	.211	.065	−1.02	.02
	Control	High	−.570*	.227	.045	−1.13	−.01
		Low	.496	.211	.065	−.02	1.02
Vänligen betygsätt filmen (1=Sämst, 10=Bäst)	High	Low	1.894*	.307	.000	1.14	2.65
		Control	.952*	.308	.010	.19	1.71
	Low	High	−1.894*	.307	.000	−2.65	−1.14
		Control	−.942*	.286	.005	−1.65	−.23
	Control	High	−.952*	.308	.010	−1.71	−.19
		Low	.942*	.286	.005	.23	1.65
Experience_Index	High	Low	1.22558*	.21565	.000	.6930	1.7582
		Control	.79177*	.21641	.002	.2573	1.3263
	Low	High	−1.22558*	.21565	.000	−1.7582	−.6930
		Control	−.43381	.20108	.101	−.9305	.0628
	Control	High	−.79177*	.21641	.002	−1.3263	−.2573
		Low	.43381	.20108	.101	−.0628	.9305
RePurchase_Index	High	Low	1.60318*	.30818	.000	.8420	2.3643
		Control	.77038*	.30927	.048	.0065	1.5342
	Low	High	−1.60318*	.30818	.000	−2.3643	−.8420
		Control	−.83280*	.28736	.017	−1.5425	−.1231
	Control	High	−.77038*	.30927	.048	−1.5342	−.0065
		Low	.83280*	.28736	.017	.1231	1.5425