Merging Science with Business: Using Psychological Knowledge to Understand and Influence Actual Purchasing Behaviour of 4,79 Million Online Shoppers
Bueters, Floor

Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen

Merging Science with Business: Using Psychological Knowledge to Understand and Influence Actual Purchasing Behaviour of 4,79 Million Online Shoppers

Date: 30/12/2022
Supervisor: Sander Nieuwenhuis
Second reader: Samarth Varma
Word count: 10121
Cognitive Psychology
Thesis Msci Applied Cognitive Psychology

Layman's abstract

Can you influence the behaviour of online shoppers by using scientific knowledge to improve websites? In this thesis, we investigated the effects of consumers' certainty concerns on consumers' purchasing behaviour. When shopping online, trust is important. Online consumers must trust the websites and the online shopping process enough to make a purchase. In other words, consumers must have enough *certainty.* We expected that certainty would be more important for desktop visitors than for mobile visitors, because desktop visitors are more dependent on analytical information that also affect certainty concerns, whereas mobile visitors are known to make more emotion-based decisions.
This is exactly what we found. Furthermore, we also investigated different phases of the customer journey (orientiation stage and purchasing stage), and different product types (fun-oriented products and functional products). Contrary to our expectations, online shoppers in the purchasing stage were not influenced more by increased certainty than shoppers in the orientation stage. Finally, and against our expectations, online shoppers who were browsing for functional products were not influenced more by increased certainty than online shoppers browsing for fun products. Currently, it is very important to do scientific research that is applicable to the behavior of actual people in the real world. With this research, we show that business and science mutually benefit when they combine their strengths and scientific knowledge is applied to a real-life setting.

Abstract

The current thesis investigates how scientific knowledge can be applied to actual consumer behaviour. The shift to e-commerce comes with many opportunities, more specifically, the amount of data that businesses collect from their visitors makes it possible for them to collect insights about their customers' behaviour. Businesses conduct research through online randomised controlled trials called A/B-tests, which allows them to test the consequences of changes. In online commerce, there is a large importance of trust and consumer certainty. The current research investigated the effects of consumers' certainty concerns on consumers' purchasing behaviour for different devices, funnel locations, and product types. As expected, our results showed that the effects of increased consumer certainty on test uplift outcome were larger for desktop devices (H1) than for mobile devices. Contrary to our expectations, we found no significant differences for different funnel locations (H2) and product types (H3) as a result of increased consumer certainty. Exploratory analyses showed no significant interaction effects between conditions. By investigating actual behaviour in real-life settings, we answer the call by the Dutch Research Council (as well as others) stating that the practical applicability of scientific knowledge should be increased. Our research shows the unique opportunity that online experimentation provides for both business and science to combine their strengths and mutually benefit from the available knowledge and expertise.

Merging Science with Business: Using Psychological Knowledge to Understand and Influence Actual Purchasing Behaviour of 4,79 Million Online Shoppers

What if you could research and analyse the behaviour of close to 5 million actual consumers purchasing actual products, driven by their own shopping needs, while spending their own hard-earned money? That's what we will be doing in this thesis.

Insights and knowledge derived from scientific research can make an important contribution to challenges in society and create business opportunities. Scientific communities such as universities and research institutes, however, tend to place more value on objectivity than applicability, and put more focus on acquiring fundamental knowledge than on conducting applied research. Nevertheless, the fundamental knowledge derived from scientific research has been a recognised contributor to industrial developments, and formed the foundation of a number of industries (e.g., IT and other technology-based industries). In recent years, however, there have been more initiatives that try to stimulate the societal impact of academic research by boosting interactions between the scientific community and social stakeholders. As such, the Dutch Research Council (NWO), responsible for funding and managing the national infrastructure of knowledge and research, included knowledge utilisation as a central topic in their research policy (The Dutch Research Council, 2020). Knowledge utilisation, as described by NWO, is the process of making societal impact through research. NWO founded three impact approaches to stimulate researchers to shift their perspective towards societal issues, and include social stakeholders throughout the research process.

Knowledge utilisation

Why do we need encouragement to improve the applicability of research, when there are entire industries that were built on the already existing knowledge? The answer lies in the methods by which research is often conducted and the applicability of these methods to real-world examples.

For example, a frequently used study design in academia is a randomised controlled trial (RCT). Experiments in the form of RCTs allow researchers to exert control over factors that are outside of the scope of the study. In a well-designed RCT, subjects are randomly assigned to one of at least two conditions. The experimental group is presented with an intervention, and the control group is presented with the exact same stimuli without the intervention. This way, significant differences between the control and the experimental group may be attributed to the intervention. Although it is unlikely for a single study to prove true causality on its own, RCTs are deemed to be a powerful tool that help reduce biases that are inherent to other study designs (Hariton & Locascio, 2018). In survey research, for example, there are risks of a number of biases, including nonresponse bias (only motivated people are answering) and biases related to self-reporting (participants answer questions about themselves more beneficially). This can result in inaccurate research findings that are not representative of the situation, and may not be reliable to apply in other contexts. Important to note is that RCTs only provide reliable assessments of cause-effect relationships between intervention and outcome if the number of participants included in the study is sufficient to ensure statistical power.

However, while the RCT study design provides accurate estimations of the intervention effects on the examined population, these findings may not be generalizable in other populations as a proper RCT design may require a lab setting to control for factors outside of the intervention. In more exact fields, such as engineering or IT, much of the research that is done is quantitative. Moreover, the application of the findings is done in conditions similar to the research conditions, inherently facilitating the application of the findings from this research. In other fields, such as social sciences, generalizability of knowledge is a bigger challenge (Lewis et al., 2003). Real-world conditions can differ strongly from the conditions that were artificially generated in the lab or research context, making it difficult to assess whether research findings have broader meaning outside of the research sample.

Consumer behaviour

One area of research in which the real-world conditions can cause several issues in the assessment of the findings' broader applicability is that of consumer behaviour. Here, contextual factors (e.g., perceived risk, time pressure) are often found to cause strong effects and sometimes even reversals in consumers' expressed preferences. To illustrate, O'Donnell and Evers (2019) found that the likelihood of one option being favoured over another was largely associated with the way in which they asked participants to express their preference. They compared two commonly used measures of preference to assess preferences between options, namely choice and willingness to pay (WTP). Expressed preferences consistently showed that the measure of willingness to pay resulted in a preference for more utilitarian goods, whereas the measure of choice resulted in a more preference for hedonic goods. These results indicate that consumer research using one of the aforementioned preference elicitation methods would generate findings that are biased towards a preference for either utilitarian or hedonic goods. Essentially, it could be argued that a truly unbiased assessment of consumer preference or purchase intention does not exist.

Aside from the impact of the selected assessment methods, can we really learn about consumers' preferences and purchase intentions from lab research? Participants taking part in a study need to come to the lab and are aware that they are being observed. Consequently, their attention is directed towards the task, and more importantly: they know that they will spend none of their own money. All of these contextual factors become increasingly important as consumers are moving away from physical stores and towards online shopping. The full journey from product orientation to completing purchases can now be done from the comfort of consumers' home environment. With that come all the different contextual factors and real-world distractions that influence the preferences and purchase intentions. As a result, the value of fundamental research findings for real (especially online) commerce are difficult to assess.

Luckily, there are opportunities to research actual consumer behaviour with real money and real purchases. Specifically, there are large online businesses already measuring and even conducting research on the visitors of their websites. Such online experiments are called A/B-tests: randomised controlled trials with small interventions conducted with real consumers in the real-world context. A/B-tests are a type of research that is similar to RCTs, but they are unmatched in the speed and ease in which they can be executed. Online experiments require businesses to make use of the web data and metrics that they already have, thus making it a low input and low risk research method. Take the example of Microsoft's Bing, where 200 concurrent A/B-tests can be running on the website at any time (Kohavi et al., 2013). With the worldwide shift to e-commerce, businesses realise that their online platforms provide them with heaps of data from their online visitors that describe the outcomes of their efforts. Large amounts of online visitors allow them to conduct experiments on their websites. Consequently, A/B-tests help guide business decisions in a data-driven manner to optimise their platforms as they generate experiment outcomes with high levels of precision. And while the practice of online experimenting evolved, businesses expanded the scope of A/B-testing beyond testing only the functionality and effectiveness of online innovations, and started utilising the research opportunities to learn about visitor behaviour (Kohavi et al., 2013).

One psychological construct that is particularly relevant for online consumer behaviour is certainty (Kim et al., 2008). Consumer certainty plays an important role in most purchasing decisions, and an even greater role in online purchasing decisions. Large technological advances in the 21st century that drove consumers towards e-commerce also allowed businesses to much more easily collect and potentially distribute individuals' personal information. As a result, online consumers are uncertain about how websites are handling their personal information, and fear that their online privacy may be violated. These kinds of online privacy concerns are found to negatively affect the usage of e-commerce platforms. Concerns can be directed towards the legitimacy of the business, safety of the payment methods, or the potential breach of users' privacy (Maseeh et al., 2021).

The current research

In this thesis, we examine to what extent current research on consumer certainty is applicable to online consumer behaviour. More specifically, we assess the effects of increased consumer certainty on purchasing behaviour for different devices (desktop and mobile), funnel locations (upper and lower), and types of product (utilitarian and hedonic).

For our research, we were able to access the experiment database of an online optimization consultancy that specialises in psychology-driven A/B-testing. The database consists of A/B-tests with alterations and hypotheses based on scientific knowledge on psychology and consumer behaviour. We collected information on one of their psychology-based optimization strategies that is often applied in the A/B-tests: increasing consumers' certainty about the product, the online process, or the organisation. The online test data consists of visitors' behavioural metrics, and test metadata containing a number of variables with contextual information for all tests. We have access to the combination of this metadata on the tests' contextual factors and insight into the intervention presented in the test. We will examine to what extent scientific knowledge on consumer certainty is applicable to online consumer behaviour.

Online experimentation

The practice of online A/B-testing is relatively new to the business world (Kohavi et al., 2013), which explains why there have not been many studies on the application of scientific research in online experimentation. Miller and Hosanagar (2021) did a meta-analysis with the aim of determining which are the most effective types of A/B-test applications. The A/B-test results in their meta-analysis were not based on psychological or scientific knowledge and the test metadata was therefore used to perform an unstructured text analysis and determine the type of alterations that were made in the test. The categories that resulted from these analyses were mostly price- and value-focused. Research on the true value of price and value differences refers to the "axiom of greed" stating that – other things equal – more of a good is preferred to less of a good (Seuntjens et al., 2015). This assumption that people always aim to maximise their outcome, however, does not take into account that perceived value is not always absolute, or rationally evaluated (Van Dijk, 2013). In summary, human behaviour is largely influenced by factors that are unrelated to absolute (monetary) value. This further emphasises the importance of considering factors outside of price or monetary value in the world of business and A/B-testing.

Currently, the majority of A/B-tests are created in a marketing- and commerce-driven way: attempts to increase websites' absolute value offer, rather than the perceived value for consumers. Increasing certainty is one of five psychological strategies used by the consultancy to increase online purchases. Interestingly, more certainty does not always lead to a higher purchase intention in consumers. Laran and Tsiros (2013) studied the effects of decreasing, rather than increasing certainty, and differentiated between cognitive and affective decisions in their research. They found contrasting results: in cognitive decisions, uncertainty negatively affected purchase intention, but in affective decisions, conversely, it increased purchase intention. The data in our database provides indicators by which we can assess the effectiveness of increased consumer certainty on purchasing decisions. The test metadata contains information on the devices on which tests were executed, in which part of the website, and the business the website belonged to.

Device

Device differences are a common occurrence in e-commerce. According to the Global UX Map report made by the digital insights platform Contentsquare (2019) mobile users currently make up the largest percentage of internet users, but it doesn't translate into mobile consumers purchasing products online. Many businesses are familiar with the different purchasing behaviour between mobile and desktop, and the phenomenon is now even referred to as the *"conversion gap"*. As it appears, a difference in shopping motivations exists for the different devices. Hedonic motivation is an important driver in the initiation and continuance of shopping on mobile devices (Luceri et al., 2022). Whereas on desktop devices there is a larger emphasis on the rational aspects of the purchase process for mobile devices this is more geared towards the emotional aspects. When we compare these findings to the findings from the 2013 study by Laran and Tsiros, the utilitarian motivation on desktop fits the cognitive decision-making process, whereas mobile shopping decisions rely more on affective processing.

Funnel

Test location is another important factor in the predicted effectiveness of increasing consumer certainty. Consumers' safety and privacy concerns are mostly concerned with the sharing of personal information, and the potential of their data being distributed and abused (Anic et al., 2019). In terms of e-commerce, this concern is especially relevant in the later stages of online shopping. In the check-out stages, often referred to as the lower end of the purchasing funnel, consumers are asked to fill in their personal details, and eventually their bank details as the payment is made. Filling in personal information and making transactions through an online banking platform is highly risk sensitive (Anic et al., 2019), therefore increasing the importance of consumer certainty in this stage.

Product type

Considering the importance of the different types of motivation (utilitarian and hedonic), and the different types of processing (cognitive and affective) for the effectiveness of consumer certainty, we also consider the different product types (utilitarian and hedonic) as a potential factor of influence. The type of product that a consumer is looking to buy is an important factor in determining which type of processing they use for their decision, or which information they are most likely to pursue. Verhagen and colleagues (2010) examined whether consumers shopping for different product types (utilitarian or hedonic) relied more on different website content elements to make their purchase decisions. They found that consumers shopping for utilitarian products relied more on certainty-increasing elements such as company information and advice content, whereas hedonic shoppers preferred a large and unique assortment of options to choose from. The metadata in our original database did not include information regarding the product type sold on the website on which the tests were done. Based on previous research on hedonic and utilitarian products (Strahilevitz & Myers, 1998), and the categorisations as described by Verhagen and colleagues (2010) we categorised the tests by classifying the included companies' websites as being either functional (utilitarian) or pleasurable (hedonic).

Hypotheses

In this thesis, we will investigate three hypotheses, to understand how increasing certainty influences online purchasing behaviour. Online purchasing behavior is operationalized by a change in uplift: the relative difference between the percentage of consumers finalising a purchase in each experimental group (Kohavi et al., 2013).

Device. The *cognitive* type of processing that is more prominent in shopping on desktop is highly reliant on actual and analytical information. In mobile shopping, decisions are more dependent on *affective* cues, such as imagery and uniqueness. Our chosen strategy of increasing consumers' perceived certainty strongly appeals to the *utilitarian* motivation that is important for desktop users. We hypothesise (H1) that for tests on desktop devices, alterations to improve consumer certainty have a larger effect on the outcome in uplift, as compared to the effect on mobile device outcomes.

Funnel. Increasing certainty about consumer security and privacy can partially mitigate consumers' privacy concerns when making purchases online. In the *upper* funnel, these concerns are not yet prominent, since the website visitor is yet to make the purchase decision. In the *lower* funnel, however, the perceived security risks are evoked, since the payment is done, and personal information is asked. Therefore, the relevance of improving visitors' perceived certainty increases in the *lower* funnel. We hypothesise (H2) that alterations to improve customer certainty have a larger effect on the outcome in uplift for tests in the lower funnel, as compared to the effect in the upper funnel.

Product type. Certainty-increasing alterations appeal to the utilitarian motivations that are prevalent in purchasing utilitarian type products. Uniqueness and surprising factors have been found to be effective for hedonic type products. We hypothesise (H3) that alterations to improve customer certainty have a larger effect on the uplift for websites selling utilitarian product types, as compared to the effects for websites selling hedonic product types.

Methods

Data Collection

The original data was collected from the database of a consultancy company specialised in online optimization of e-commerce websites by performing A/B-tests. The company works with large e-commerce businesses active in the Netherlands, Belgium, and Germany. The businesses' websites have a large number of daily users, which enables them to analyse the data of large groups of participants in a short period of time.

Website visitors participate in these tests if they consent by allowing specific cookies to be placed. In A/B-tests, visitors are randomly assigned to either the original version of the website (the A variant), or a new version of the website (the B variant). The new version of the website is a variation of the original website in which small alterations have been made to the original elements.

A power analysis determines the duration of an A/B-test before running the test. Once the duration of the test has been completed, data analysts compare the results of both groups by performing statistical tests comparing the control group and the experimental group for the metrics targeted with the variation in the test. Through this process, businesses are able to frequently perform randomised controlled trials that provide valuable insights, and require relatively small financial and time investments.

Online measurements

Visitor data is measured on websites to support the correct functionality of essential elements, such as the retention of visitors' actions that lead to the next step of the website towards a potential purchase. The owner of the website has access to this confidential visitor data, and is not allowed to share visitor data with third parties. In countries that are part of the European Union, the General Data Protection Regulation (GDPR) states that websites cannot directly transfer their visitor data to a third-party technology without having the visitors' consent.

In order to do A/B-testing, however, companies need to implement a third-party technology. The technology enables website owners to randomly divide visitors into control and experimental groups, and to measure each group separately. Compliant with the GDPR, companies inform their visitors about their online policy, and ask their consent for the data to be transferred to a third-party technology.

The data consists of metrics that indicate online consumer behaviour and purchases, as well as contextual information such as the device on which they visited the website, and functionality metrics of the website such as loading time. Importantly, in compliance with the GDPR, all visitor data was anonymized, and can not be retraced to individual website visitors. The database used in this thesis consisted of the aggregated visitor data of a combined total of 65,307,715 website visitors.

Psychological optimisation strategies

Our original database consists of the results of 1404 psychology-driven A/B-tests performed on the e-commerce platforms of 9 companies between October 2018 and June 2022. The A/B-tests were performed using the services of a consultancy company based in The Netherlands, called Online Dialogue. This consultancy specialises in the optimization of online platforms, in which psychological knowledge is applied as the foundation of all A/B-tests and business strategy advice. Online Dialogue consultants that have a higher education in psychology identify the issues or opportunities that are exposed in the available data on e-commerce platforms. They identify the psychological phenomenon at the core of the issue, and subsequently create optimizations to be tested. Each A/B-test has a measurable, and directed hypothesis that is created to confirm or reject the optimizations.

Behavioural Online Optimization Method. The consultants at Online Dialogue that specialise in psychology created a framework to categorise the type of alterations that are used in A/B-tests. The *Behavioural Online Optimization Method* (BOOM) consists of five different psychological strategies based on psychological phenomena that have been well-established in scientific literature. The strategies are first used as a means to find and recognise potential issues and bottle-necks and to then optimise according to the identified customer needs, motivations, and abilities.

Consultants draw from a number of information sources (e.g. the available webdata, scientific literature, and previous experiments) to examine potential problems and opportunities at the start of the A/B-testing process. A psychologist then chooses the BOOM strategy that is most applicable to the situation, and forms a hypothesis for the expected effect of the alteration. The hypothesis is then tested in one or multiple A/B-tests. The classification of the experiments through the BOOM framework leads to broader insights on customer behaviour and needs. In this framework, the online optimization strategies are divided into five categories as summarised in Table 1.

Table 1

Psychological optimization strategies.

BOOM strategy	**Application**
Ability	Increase ease of the desired behaviour
Attention	Direct attention towards the desired element
Motivation	Appeal to users' existing needs and goals
Certainty	Increase confidence about the product, the process, or the organisation
Choice Architecture	Influence choice by changing how options are presented

Participants and Interventions

Participants

All participants were users of the websites on which the website owners ran controlled experiments. All websites were compliant with the General Data Protection Regulation (GDPR) to ensure data privacy for users. After users agreed to the website's cookie policy, they were randomly assigned to the control group or the variant group. The GDPR does not allow for any personal information to be transferred to the A/B-testing database that we used. Therefore, no personal data on participants, but only test data is described.

Manipulation checks

In order to confirm that visitors have seen the altered element when it is not directly visible upon visiting a webpage (i.e., it is only visible when certain buttons are clicked, or visitors are scrolling below a certain point) a *viewport* can be implemented. The *viewport* measures whether the element that was altered was visible on the device of website visitors, in both the control group and the experimental group.

Moreover, the A/B-test technology ensures that visitors that are assigned to the control group or the experimental group will remain in that group for the duration of the test. Consequently, visitors are presented with the same version of the website each time they visit. As such, differences in outcome between the groups as a result of the alteration are measured throughout the duration of the test, and not merely during a single website visit.

Conditions

E-commerce businesses randomly divide their visitors into two groups by integrating a third party technology on their website. When users have agreed to the terms and conditions applicable to the website, a control group is presented with the original version of the website, and a test group is presented with the same website on which an element has been altered. For both user groups, the metrics that provide insights into user behaviour are registered and documented by the A/B-test technology. Additional user information (e.g. used device or location) is registered, depending on the website and goals of the e-commerce business.

Procedure

Website users are not aware which of the versions of the website they are presented with. Website users are visiting websites of their own choosing, and shopping online as they regularly would.

E-commerce businesses determine the duration of an A/B-test before presenting it to the users. With a statistical power of 80%, for each of the webpages the minimum duration of the tests are determined with a calculation of the minimum detectable effect (MDE) to be able to confirm that the alteration made a statistical difference.

Dependent variables

In this research, we consider the metrics related to consumer behaviour. Conversion rate and Click Through rate are the most direct indicators of online consumer behaviour and visitors' progress towards an eventual purchase throughout the customer journey.

Conversion rate	The percentage of visitors in the control group or the variant group who have completed a transaction.
Click-through rate	The percentage of visitors in the control group or the variant group who have proceeded to the next step in the funnel.

Uplift outcome

Conversion rate and Click-through rate for both the control group and the experimental group are expressed as a percentage, indicating the portion of visitors in the group performing the specified action. In order to determine the effect of the alteration as compared to the control, the outcomes of the two groups are compared. The relative difference between the two group outcomes is called the uplift of an A/B-test. This difference is expressed as a percentage, and the outcome can be positive or negative. We analysed the uplift data for the A/B-experiments in this thesis. Consequently, any visitor data that is used in this thesis reflects the customer behaviour on the company's website in its entirety, and not interactions of individual visitors or individual products.

Independent variables

In the original database various kinds of metadata on A/B-tests is included. In order to test our hypotheses, data on device, funnel, and product type was collected for further analysis.

Device. There are two categories for online devices: Mobile, Desktop. A/B-test results show a separate *uplift* of the metrics comparing the control group with the variant group.

Funnel. In e-commerce, and A/B-testing research, the customer journey on a website is referred to as a funnel. The elements and webpages and their purpose on a website can be categorised through various frameworks (Miller & Hosanagar, 2021) defining the stages in the funnel. Our database includes data on the webpage (e.g., homepage or check-out) on which the alteration in the A/B-test was made.

In order to research the difference between decision making and finalising a purchase, we divide the funnel into two categories: the stage *before* the decision to purchase, and the stage *after* the decision to purchase was made. On e-commerce websites, the homepage, product page, and other pages before the shopping cart enable visitors to browse, search for products, and acquire product information necessary to make their decision. Once the decision to purchase is made, visitors continue the customer journey by visiting the shopping cart, and check-out page(s) on which they can finalise their purchase.

Pages prior to visitors adding products to, or clicking on the *shopping cart* (pre-decision) are categorised as being *Upper Funnel*. All web pages from the *shopping cart* to the confirmation of finalising the purchase (post-decision) were categorised as being *lower funnel*.

Product type. The database includes information on the nature of the products that are sold on the company's website. We use the categorisation of products defined by Strahilevitz and Myers in 1998 that makes a distinction between fun and functional purposes for products, categorised as utilitarian and hedonic. Products that are a necessity (e.g., office supplies, food products, or laundry detergent), and goal-oriented were described by Strahilevitz and Myers as *Utilitarian*. Products that are purchased with the purpose of entertainment, and pleasure-based were categorised as *Hedonic* (e.g., an exclusive cologne or a luxury vacation).

Inclusion criteria

The following criteria were applied for selecting suitable tests for further analysis:

- Test had the main goal of directly optimising the customer journey toward a conversion. Included were tests with the following key performance indicators (KPI):
  - *Conversion rate* (CR)
  - *Click Through rate* to next page (CTR)
  - *Logins*
- The following list of test KPI's did not have the main goal of directly optimising the customer journey toward a conversion:
  - *Added to favourites*. Favourite lists are frequently used as wish lists, and therefore not directly related to any conversions in the session.
  - *Customer Service contact*. When tests are done to improve metrics related to customer service, the outcome is not directly related to the number of conversions.
  - *Event clicks*. KPIs related to specific Event clicks can vary greatly in functionality, and are most frequently unrelated to the next step in the customer journey.
  - *Bounce* and *Click Through* to an external website. These metrics refer to the number of people that leave the website either directly after landing (Bounce) or through a link directed to another website.
  - *Account registered*. The number of registered accounts does not inherently relate to the customer journey of the session that is undertaken.
- Test was done on both Mobile and Desktop devices.

- Test alterations were made in either *Lower Funnel* or *Upper Funnel*. This ensures that the effect is indeed related to the location of the alteration, as opposed to elements that are shown throughout the website.
- Test was done on a website selling physical products, excluded were websites selling services or online products.
- Test had no more than two experiment groups and the visitor samples in the control group and the variant group were equal in size. The technology ensures that the control group and the experimental group have matched sample sizes. The visitor groups are randomly divided. When the sample sizes in a test result in unequal group sizes, the random division of visitors can not be ensured, and these tests will be excluded from our dataset.
- Test had a minimum of 200 visitors per device, and per group (both the control group and the experimental group).
- Test had a minimum of 50 conversions per device in the control group. The *uplift* outcome of a test is calculated in proportion to the number of conversions in the control group. A minimum of 50 transactions excludes extreme outcomes that are a disproportionate indicator of the impact of the alteration in the experimental group.

After the application of the exclusion criteria, there were a total of 608 tests remaining for further analysis.

Selecting a psychological strategy

The goal of this thesis was to examine the real-world effectiveness of scientific knowledge on consumer psychology, and investigate potential boundary conditions. We used the outcome data of A/B-tests performed on websites with alterations based on psychological knowledge of five different phenomena. A prerequisite for forming hypotheses on the effects of the psychological strategies on the tests' uplift outcome, and for examining the hypothesised effects based on the psychological literature, was to choose one of the five psychological strategies for further analysis.

The five psychological strategies in our database are hypothesised to have opposing effects in the context of online consumer behaviour. In order to test our hypotheses, we will evaluate the strategies and select the most suitable psychological strategy to include for further analysis. We formulated the following criteria to evaluate the characteristics of the strategies and determine the strategy most suitable for further analysis.

Missing BOOM data. The original database contained a number of tests with no BOOM categorisation, or with multiple BOOM strategies. A total number of 16 tests were excluded on the basis of not being categorised with a single BOOM strategy. After the exclusion of these tests, the remaining number of tests was 592.

Directionality of the strategy. To determine whether the psychological strategy was suitable for further analysis, we started by assessing the alignability of different applications used within the strategy. We found that the strategies *Choice Architecture* and *Motivation* can correctly be applied in more than one direction. For example, two tests in the same category can have directly opposite applications (i.e., *Choice Architecture* can be applied as ‘removing an option’, but also as ‘adding an option’). There is no metadata in the database to further examine the direction of these strategies, therefore *Choice Architecture* and *Motivation* are not considered for further analysis.

Variance of uplift outcomes. We formulated hypotheses for the three remaining optimization strategies eligible for further analysis: *Ability, Attention*, and *Certainty*. The hypotheses state the direction of the expected differences in uplift outcome of the tests between conditions in device, funnel and product type.

The selection of a strategy was based on the results of preliminary analyses to assess the significance of the differences in uplift outcomes between conditions. In order to test our main hypotheses in further analyses, we aim to choose a strategy with sufficient variance.

Preliminary analysis

After applying the exclusion criteria, the remaining number of tests per psychological strategy is summarised in table 2.

Table 2

Number of tests results per strategy after exclusion

Optimisation strategy	Number of tests
Ability	188
Attention	152
Certainty	112

Tests for normality

The first step in the preliminary analysis was to test the uplift distribution of the remaining strategies for normality in the conditions for device, funnel, and product type. Considering the small sample size of test uplifts per strategy, we used Shapiro-Wilk's test to assess the normality of distributions of test uplifts. We found that test uplift outcomes were not normally distributed for each of the strategies.

We used non-parametric Mann-Whitney tests to assess the significance of differences between conditions, in order to select the strategy with sufficient variance.

Ability. A Mann-Whitney test for the uplift outcomes of Ability tests indicated that there were no statistically significant differences between device conditions, funnel conditions, and product type conditions.

Attention. A Mann-Whitney test for the uplift outcomes of Attention tests indicated that the difference between device conditions, funnel conditions, and product type conditions was not statistically significant.

Certainty. A Mann-Whitney test for the uplift outcomes of Certainty tests indicated that the difference between device conditions was statistically significant. The difference in uplift outcomes between funnel conditions, and product type conditions was not statistically significant.

Analyses plan

Based on these findings, we included tests focused on increasing certainty for online visitors in our further analyses. A total of 112 tests were included in the dataset for further analysis. We used a 2 (device: mobile vs desktop) x 2 (funnel: upper vs lower) x 2 (product type: hedonic vs utilitarian) between-subjects design with uplift as dependent variable, and device, funnel and product type as independent variables.

Results

Quality assurance

The dataset included in further analysis consisted of the uplift outcomes of a total of 112 tests. The combined total number of participants in all 112 tests was 4,791,653 with the average number of participants per test being 85,565. The distribution of the number of participants per test is presented in Table 3.

Table 3

Distribution of number of participants included per test

Participants per test	N
800 - 10000	28
10000 - 30000	28
30000 - 50000	20
50000 - 200000	18
200000 +	18

Outliers

We screened the test uplift results in our final dataset for outliers. Outliers were defined as tests with an uplift outcome further than ± 3 standard deviations away from the mean (M = .667, SD = 3.69). A total of two test uplift outcomes out of the original 112 tests fell outside of the defined outlier range. However, upon further investigation, both test uplift results (one for mobile, one for desktop) did not stand out in the context of the original database. Moreover, the direction of the results were consistent with our expectations. By not removing the two tests from the dataset, we prevented a loss of valuable information, considering the small sample size of the dataset.

In order to assess the validity of the outcomes of the two tests that were defined as outliers in our dataset, we performed an additional outlier analysis comparing the two tests to the original database before exclusion. This outlier analysis showed that the test outcomes that were outliers in our final dataset were not extreme compared to the rest of the data set, and validated the inclusion of the tests in our final analysis.

Data views

We examined three independent variables (device, funnel, and product type) within the *Certainty* data-set. The distribution of the two conditions for each independent variable (device: desktop and mobile, funnel: upper funnel and lower funnel, product type: hedonic and utilitarian) is displayed in Table 4.

Table 4

Distribution of conditions in Certainty tests

	Device		*Funnel*		*Product type*
Group	N	Group	N	Group	N
Desktop	56	Lower	74	Hedonic	26
Mobile	56	Upper	38	Utilitarian	86

Descriptive data

We compared the test uplifts for the different conditions in device, funnel, and product type to test our main hypotheses. Table 5 displays the means and standard deviations of the two conditions in device, funnel, and product type.

Table 5

Test descriptives (means, SD, std. error)

	Device		*Funnel*		*Product type*	
Group	Desktop	Mobile	Upper	Lower	Hedonic	Utilitarian
Test uplift mean %	+ 1.36	-0.01	+ 0.89	+ 0.57	+ 0.54	+ 0.72
Std. Deviation	3.30	3.97	4.51	3.23	2.81	3.94
Minimum	-4.6	-14.8	-14.8	-4.9	-3.5	-14.8
Maximum	13.5	10.8	8.8	13.5	8.8	13.5

Test for normality

To further analyse the differences between conditions, we first assessed the assumption of normality of the test uplifts for device, funnel and product type conditions using Shapiro-Wilk’s test, which was selected due to the small size of the samples. The assessment of uplift distributions in all conditions showed a significant departure of normality (significance for all conditions were below the cutoff value of p-value <.05, see appendix A).

We performed non-parametric Mann-Whitney tests as a result of the non-normally distributed uplifts to examine the main effects of device (H1), funnel (H2), and product type (H3) on outcome in test uplift. We performed additional exploratory analyses to explore potential two-way and three-way interactions between device, funnel, and product type. A univariate ANOVA was performed to examine the exploratory analyses.

Confirmatory analyses

Device

Website visitors included in the data-set use one of two devices (desktop or mobile) to do online shopping. In order to examine the first hypothesis, we performed a Mann-Whitney test for independent samples to compare test uplifts in the two device conditions 'desktop' and 'mobile'.

A Mann-Whitney test indicated that this difference was statistically significant, $U(N_{\text{Desktop}}= 56, N_{\text{Mobile}}= 56,) = 1228.50$, $z = -1.98$, $p = .048$. Test uplift outcomes for desktop users (*Mdn* = 1.1) were higher than those of mobile users (*Mdn* = -0.2).

These results support our first hypothesis that on desktop devices, alterations to improve customer certainty have a larger effect on the outcome in *uplift*, as compared to the effect on mobile device outcomes.

Funnel

In order to examine the second hypothesis, we performed a Mann-Whitney test for independent samples to compare the test uplift in the two funnel conditions 'upper' and 'lower'.

Contrary to our expectations, a Mann-Whitney test indicated that this difference was not statistically significant, $U(N_{\text{Upper}}= 38, N_{\text{Lower}}= 74,) = 1559.50$, $z = .943$, $p = .345$. These results did not support our second hypothesis.

Product type

In order to examine the third hypothesis, we performed a Mann-Whitney test for independent samples to compare the test uplift in the two product type conditions 'hedonic' and 'utilitarian'.

Contrary to our expectations, a Mann-Whitney test indicated that this difference was not statistically significant, $U(N_{\text{Hedonic}}= 26, N_{\text{Utilitarian}}= 86,) = 1190.00$, $z = .496$, $p = .620$. These results did not support our third hypothesis.

Exploratory analyses

We performed exploratory analyses to examine potential two-way and three-way interaction effects. A three-way ANOVA was performed to explore interaction effects on test uplift between device, funnel and product type.

Two-way interactions

A three-way ANOVA was done on the dataset consisting of 112 test results to explore the two-way interactions between device and funnel, device and product type, and funnel and product type on test uplift. There was no significant two-way interaction between device and funnel, $F(1, 104) = 1.69$, $p = .197$, device and product type, $F(1, 104) = 0.32$, $p = .574$, funnel and product type, $F(1, 104) = 0.14$, $p = .705$.

Three-way interactions

A three-way ANOVA was performed to explore a three-way interaction effect on test uplift between device, funnel and product type. There was no significant three-way interaction between device, funnel and product type, $F(1, 104) = 0.07$, $p = .786$.

Discussion

In this thesis, we investigated to what extent research on consumer certainty is applicable to real-life online consumer behaviour. More specifically, we assessed the effects of increased certainty on consumers' purchasing behaviour for different devices, funnel locations, and product types. As expected, our results showed that the effects of increased consumer certainty on test uplift outcome were larger for desktop devices (H1) than for mobile devices. Contrary to our expectations, we found no significant differences for different funnel locations (H2) and product types (H3) as a result of increased consumer certainty. Furthermore, exploratory analyses showed no significant two-way or three-way interaction effects between device, funnel location, and product type.

In the current study, we investigate actual real-life purchasing behaviour of real consumers on real websites. More specifically, behaviour from 4.79 million visitors is included in the database we analysed. By investigating actual behaviour in real-life settings, we answer the call by the Dutch Research Council (as well as others) stating that the practical applicability of scientific knowledge should be increased (The Dutch Research Council, 2020). In this thesis, we investigated the effects of increasing certainty on online consumer behaviour. By doing so, we build on, and extend, the literature on online consumer behaviour in important ways.

Certainty

First of all, the finding that the effect of increased certainty on purchasing behaviour is larger for desktop devices than for mobile devices, is in line with previous research. This difference has been argued to exist because key drivers of mobile commerce are more *hedonic* by nature (Luceri et al., 2022). Mobile purchases are based on affective decisions rather than cognitive decisions, thus increasing the influence of emotional drivers of behaviour. Convenience, availability and whether the purchasing experience was pleasant all contribute to a positive affect in mobile consumers. Desktop consumers, on the other

hand, are more susceptible to rational drivers of behaviour, such as increased certainty, because they are more likely to make cognitive purchasing decisions.
The innovative nature of mobile technology, and the relative newness of m-commerce may particularly attract customers to whom these affective factors (curiosity, innovativeness) are appealing. This explains why we did not find an uplift in purchasing behaviour when certainty was increased for mobile customers.

Although we did not find an interaction between product type and device, it would be interesting to examine whether the hedonic focus of mobile shoppers is device-related or simply an indication of mobile consumers' already existing preferences. In the current study, this focus seems to be device related. This may, however, also have to do with the availability of different types of devices. More specifically, around half of the current internet traffic in Europe still uses desktop devices, while in Africa, around two thirds of the traffic comes from mobile devices, and in India mobile devices make up almost three quarters of the digital traffic (Kinsta, 2022). Ashraf and colleagues (2021) compared the use of m-commerce in developed countries to developing countries. These researchers found that m-commerce in developing countries is much more utility-driven, since for these consumers desktop devices may not be as widely available and both utility and hedonic driven purchases are made on mobile devices. Our research examined the purchasing behaviour of European consumers on European websites. We found support for our expectation that providing a utilitarian motive such as certainty is more effective on desktop devices. In future research, device preference and product types can provide more insight into the interactions between hedonic and utilitarian motivations and products.

Insights, such as from the current study, into behavioural differences between mobile and desktop customers is directly relevant to organisations. Businesses' user bases are increasingly consisting of mobile users, as the share of overall mobile shopping has increased rapidly (Thongpapanl et al., 2018). However, the conversion rates in m-commerce have not been increasing accordingly, leading to a conversion gap between desktop and mobile devices. Currently, solutions to close the conversion gap between devices are mostly sought for in the physical differences between devices such as screen size and screen touchability (Kim & Sundar, 2016). In the current study, we show how different psychological needs, such as the need for certainty, are relevant for influencing purchasing behaviour. More specifically, integrating the results from our study with earlier work by Ashraf and colleagues (2021) indicates that only investigating readily available characteristics (such as device type) is not enough to understand consumer behaviour: One should also aim to understand underlying psychological processes. All experiments in this thesis focused on the effects of increasing certainty; on one of five psychological constructs used as strategies in online experimentation by trained psychologists. To our knowledge, other research on A/B-testing did not base their test alterations on psychological theories, models, or behavioural science (see, e.g., Miller et al., 2021). Therefore, the current study has added value in terms of the included A/B-tests, and the broader perspective it examines regarding current certainty literature and the practical online implementation.

The current research has demonstrated that psychological theories on consumer behaviour are applicable in online consumer behaviour. We also demonstrated that the wealth of data generated by online businesses has scientific value, not only to validate previous findings, but to perform reliable research. Business platforms can provide the environmental and contextual factors that give more insight into the applicability and relevance of scientific research.

Stages of the customer journey

Second, this thesis extends the literature on the effect of stages in the customer journey on purchasing behaviour. More specifically, in the current study, we included *all* pages before the cart, rather than excluding important pages such as the homepage. We excluded tests that took place on sitewide elements (such as the website's header) in order to avoid creating a third category of tests that was not specifically related to a customer journey stage. As a consequence of including all pre-purchase related pages, our results show a large variance in upper funnel results. This is in contrast to earlier A/B-test studies, as described in a recent meta-analysis (Miller & Hosanagar, 2021), where the homepage was excluded. However, excluding pages such as the homepage from the funnels makes it difficult to investigate the complete customer journey, which impedes applicability and generalisability of the findings. The distinction between different stages of the online customer journey specifically has not yet been widely researched. However, in consumer behaviour research, on- and offline, the distinction between stages is commonly between the orientation process and the final purchase (Tueanrat et al., 2021). These stages are often referred to as the pre-purchase stage, the purchase stage, and the post-purchase stage. In e-commerce, the first two stages are deemed to be the most important, for businesses and customers alike. Instead of looking at the customer journey from a merely web-based perspective, it is valuable to take scientific research into account. Real-life commerce does not have a homepage. Therefore, research gives no indications of the homepage being separate from the other stages of the customer journey. Currently, it remains unclear as to what the hypothesised effects of increased certainty on consumers would be when tested on the homepage alone.

Although we included all relevant pages in either the upper or lower funnel, our results indicate it would have been even better to divide the upper funnel into two sections rather than one. Previous research made a distinction within the orientation stage between a *broader* and a more *directed* orientation stage, also referred to as the *Need Recognition* and the *Information Search and Evaluation* stage (Santana et al., 2020). This would have likely reduced the large variance we observed in the orientation stage and allowed for more accurate comparisons between customer stages and the effects of certainty. However, we had a limited number of tests in each of the funnel conditions. Dividing this into more separate categories was therefore undesirable, because it would make it difficult to draw strong conclusions.

Utilitarian vs hedonic products

In the current study, we divided products in hedonic or utilitarian, based on the website’s main focus. By doing so, we were able to investigate whether increased certainty is more relevant in purchasing behaviour of hedonic or utilitarian products. However, we found no effect of product type on purchasing behaviour. This may have to do with the complexities of current websites. The categorisation of hedonic and utilitarian product types in the current study was done for websites with a relatively homogenous catalogue. This distinction, however, may not have been so clear-cut. A utilitarian product can be sold on a website with a lot of imagery and points to be collected through a loyalty programme playing into customers' hedonic motivation. And hedonic products sold on e-commerce platforms require a checkout process to finalise the purchase, which is primarily utilitarian in nature. Future research could distinguish between more hedonic and utilitarian products within a catalogue, as frequently the two types are sold on the same website.

Limitations

In A/B-testing, it is common to investigate a single outcome variable of an experiment (i.e., the uplift) rather than comparing data from individual participants (as is more common in social scientific research). This resulted in a relatively small number of test results (n = 112) that were included in our analyses. However, even though the number of tests is relatively small, we analysed the behaviour of nearly 5 million actual online consumers.

Each test outcome was the result of a different application to increase consumer certainty. For instance, on one test, certainty was increased by providing more visual feedback to confirm the actions taken, whereas in another test, it was increased by removing unnecessarily large amounts of information that could induce doubt in consumers. We chose to weigh each test result equally in our analyses, regardless of the number of participants included in the test. Larger tests could include a number of participants up to a factor of >100 than smaller tests. Weighing all tests equally ensured that the alterations done in tests with smaller participant numbers would not be neglected in our analyses. This prevented a loss of valuable information, since the alterations made in the smaller tests were equally as insightful as those done in larger tests. Due to the small sample size of test results in our analyses, the results of the tests with smaller participant numbers could have made a relatively large impact as compared to the number of consumers that were measured.

A second limitation is related to pre-existing differences between conversion rates. All test results were the outcome in uplift of a comparison between the conversion rate of a control group and the conversion rate of an experimental group. There are preexisting differences in conversion between websites, devices and funnel locations. In this study, we did not consider the starting conversion rate of each test. The test results are expressed as uplift, which is relative to the conversion rate in the control group to ensure an equal evaluation of the test impact.

Aside from the criterion that tests were done on both desktop and mobile devices, we included no specific criteria for the conditions of funnel location and product type. This led to an unequal number of tests for funnel and product type conditions. Our sample included more tests for the lower funnel condition, as compared to the upper funnel condition, and more tests for the utilitarian product type condition than the hedonic product type condition. Based on the psychological knowledge of the consultants performing the tests, this was what would be expected for the application of increasing certainty. For both lower funnel tests and for utilitarian product type tests, increasing consumer certainty is expected to have greater potential than in upper funnel and hedonic product type tests. The different number of tests per funnel and product type condition, however, does appear to have effects on our data analyses results. Especially for the product type conditions, as the division between the two groups is the one with the largest difference (86 tests for utilitarian, and 26 tests for hedonic product type), and it has the least significant effect on the uplift outcome.

Conclusion

Online experimentation provides a unique opportunity for both business and science to combine their strengths and mutually benefit from the available knowledge and expertise. In this study we did just that, to show the benefits of conducting randomised controlled trials in real life. Generalizability being a major challenge in the more qualitative and social areas of science (Lewis, 2003), the ability to conduct research on the actual context in which the research applies adds great relevance and validity to the research. The relatively large amounts of (often unused) data resources and funds available to businesses complement the wealth of knowledge available in scientific research.

Businesses, on the other hand, should aim to benefit from the potential that their A/B-testing practices offer. In order to fully leverage the insights that their data and research methods offer, there is a wealth of scientific knowledge still to be discovered. Consumer research and current knowledge on human behaviour offer a much broader perspective on the opportunities for businesses in e-commerce. The combination of the available behavioural metrics and metadata from online experimentation and scientific knowledge can incentivise businesses to shift their focus from merely functional to human-centred online experimentation.

References

Anic, I.-D., Škare, V., & Kursan Milaković, I. (2019). The determinants and effects of online privacy concerns in the context of e-commerce. *Electronic Commerce Research and Applications*, *36*, 100868. https://doi.org/10.1016/j.elerap.2019.100868

Contentsquare. (2019). *The Global UX Map. Localizing the eCommerce experience for international audiences.* https://go.contentsquare.com/hubfs/US_2019_05_Reports_UX_Global%20Map_A5_WEB.pdf?hsCtaTracking=205d772a-3375-48b5-96df-b1eb51222bcb%7Cefefee5e-5112-4d1c-a56a-1068483048d0

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(13), 1716–1716. https://doi.org/10.1111/1471-0528.15199

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision support systems*, *44*(2), 544-564.

Kim, K. J., & Sundar, S. S. (2016). Mobile Persuasion: Can Screen Size and Presentation Mode Make a Difference to Trust?: Heuristic Versus Systematic Nature of Mobile Persuasion. *Human Communication Research*, *42*(1), 45–70. https://doi.org/10.1111/hcre.12064

Kinsta. (2022). *Mobile vs. Desktop Market Share and Usage Statistics In 2022.* https://kinsta.com/mobile-vs-desktop-market-share/

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., & Pohlmann, N. (2013). Online controlled experiments at large scale. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1168–1176. https://doi.org/10.1145/2487575.2488217

Laran, J., & Tsiros, M. (2013). An Investigation of the Effectiveness of Uncertainty in

Marketing Promotions Involving Free Gifts. *Journal of Marketing*, *77*(2), 112–123. https://doi.org/10.1509/jm.11.0255

Lewis, J., Ritchie, J., Ormston, R., & Morrell, G. (2003). *Generalising from qualitative research. Qualitative research practice: A guide for social science students and researchers, 2,* 347-362.

Luceri, B., (Tammo) Bijmolt, T. H. A., Bellini, S., & Aiolfi, S. (2022). What drives consumers to shop on mobile devices? Insights from a Meta-Analysis. *Journal of Retailing*, *98*(1), 178–196. https://doi.org/10.1016/j.jretai.2022.02.002

Maseeh, H. I., Jebarajakirthy, C., Pentecost, R., Arli, D., Weaven, S., & Ashaduzzaman, Md. (2021). Privacy concerns in e-commerce: A multilevel meta-analysis. *Psychology & Marketing*, *38*(10), 1779–1798. https://doi.org/10.1002/mar.21493

Miller, A. P., & Hosanagar, K. (2021). *An Empirical Meta-analysis of E-commerce A/B Testing Strategies*. 36.

O'Donnell, M., & Evers, E. R. K. (2019). Preference Reversals in Willingness to Pay and Choice. *Journal of Consumer Research*, *45*(6), 1315–1330. https://doi.org/10.1093/jcr/ucy052

Santana, S., Thomas, M., & Morwitz, V. G. (2020). The Role of Numbers in the Customer Journey. *Journal of Retailing*, *96*(1), 138–154. https://doi.org/10.1016/j.jretai.2019.09.005

Seuntjens, T. G., Zeelenberg, M., van de Ven, N., & Breugelmans, S. M. (2015). Dispositional greed. *Journal of Personality and Social Psychology*, *108*(6), 917–933. https://doi.org/10.1037/pspp0000031

Strahilevitz, M., & Myers, J. G. (1998). Donations to Charity as Purchase Incentives: How Well They Work May Depend on What You Are Trying to Sell. *Journal of Consumer Research*, *24*(4), 434–446. https://doi.org/10.1086/209519

The Dutch Research Council (NWO). (2020, February). *Knowledge utilisation*. https://www.nwo.nl/en/knowledge-utilisation

Thongpapanl, N. (Tek), Ashraf, A. R., Lapa, L., & Venkatesh, V. (2018). Differential Effects of Customers' Regulatory Fit on Trust, Perceived Value, and M-Commerce Use among Developing and Developed Countries. *Journal of International Marketing*, *26*(3), 22–44. https://doi.org/10.1509/jim.17.0129

Tueanrat, Y., Papagiannidis, S., & Alamanos, E. (2021). Going on a journey: A review of the customer journey literature. *Journal of Business Research*, *125*, 336–353. https://doi.org/10.1016/j.jbusres.2020.12.028

Van Dijk, E. (2013). *The Economics of Prosocial Behavior*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399813.013.015

Verhagen, T., Boter, J., & Adelaar, T. (2010). The Effect of Product Type on Consumer Preferences for Website Content Elements: An Empirical Study. *Journal of Computer-Mediated Communication*, *16*(1), 139–170. https://doi.org/10.1111/j.1083-6101.2010.01536.x

Appendix A

Tests for normality

Table A1

Test results of Shapiro-Wilk test for normality in device, funnel and product type conditions

Condition	Statistic	df	Sig.
Device			
Desktop	.950	56	.021
Mobile	.931	56	.003
Funnel			
Upper	.914	74	<.001
Lower	.928	38	.018
Product type			
Hedonic	.885	26	.007
Utilitarian	.953	86	.003

Three-way ANOVA

Table A2

Descriptive statistics of three-way ANOVA for interaction effect of device, funnel and product type on test uplift.

Condition			Measures		
Device	Funnel	Product type	*Mean %*	*Std. Deviation*	*N*
Desktop	Lower	Hedonic	.08	.614	5
		Utilitarian	1.00	3.528	32
		Total	.88	3.296	37
	Upper	Hedonic	1.88	2.910	8
		Utilitarian	2.61	3.450	11
		Total	2.30	3.169	19
	Total	Hedonic	1.19	2.428	13
		Utilitarian	1.414	3.5386	43
		Total	1.361	3.2950	56
Mobile	Lower	Hedonic	-.100	1.6718	5
		Utilitarian	.313	3.3616	32
		Total	.257	3.1721	37
	Upper	Hedonic	-.100	3.8767	8
		Utilitarian	-.827	6.2434	11
		Total	-.521	5.2570	19
	Total	Hedonic	-.100	3.1142	13
		Utilitarian	.021	4.2279	43
		Total	-.007	3.9710	56

Glossary

Test	An online Randomised Controlled Trial with the aim of optimising the online platform on which it is conducted. In e-commerce this is referred to as an *A/B-test*.
Participant	Website visitor that is part of the experiment, either in the control group or in the experimental group.
E-commerce	The buying and selling of goods and services through the use of online platforms such as (mobile) websites and apps.
Conversion rate	The percentage of visitors in the control group or the experimental group who have completed a transaction.
Alteration	The intervention that is done to create a variant of a website, which is often the manipulation of an element on one page on the website.
Variant	The version of a website that has alterations in page elements or other individual aspects of the website. Half of the website visitors in a test are presented with the variant, the other half are presented with the original version of the website.
Session	A single and continuous visit of a website user. Each time a visitor returns to the website it is registered as a new session.
KPI	*Key Performance Indicator*. Before the start of an A/B-test the main goal is determined for which metric is the most important to improve.

Uplift The difference in conversion rate between the control group and the experimental.

Device The means by which a website is visited, on mobile phone or desktop.

Funnel The customer journey throughout the sales process, offline and on an e-commerce platform is referred to as a funnel. The orientation process is represented in the top of the funnel (*upper funnel*) and in the bottom of the funnel (*lower funnel*) the purchase is finalised.

Product type Refers to the product that is offered on the website, and whether the purpose of the product is functionality-based (*utilitarian*) or pleasure-based (*hedonic*).