MULTI-ATTRIBUTE CONSUMER CHOICE AND DECISION CONFLICT: A PROCESS TRACING STUDY

Master's Thesis
Isa Korhonen
Aalto University School of Business
Information & Service Management
Spring 2023

Aalto University
School of Business

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

Abstract of master's thesis

Author Isa Korhonen

Title of thesis Multi-attribute consumer choice and decision conflict: A process tracing study

Degree Master of Science in Economics and Business Administration

Degree programme Information & Service Management

Thesis advisor(s) Ilkka Leppänen

Year of approval 2023 **Number of pages** 44 **Language** English

Abstract

Decision making has been studied throughout decades and even centuries, and along that journey, also the research methods to study it have been developing. A few of the current popular methods include eye-tracking and mouse movement tracking. However, both of these entail some limitations or impracticalities that make the method inflexible and/or expensive. Therefore, room for new methods exist. Especially tracing scrolling movements instead of mouse movements, as was tested in this research, could be an interesting direction for process tracing in order for the field to stay current in today's age of mobile online shopping.

In this thesis, a quantitative empirical study was done with an open-source platform oTree that was used to create a browser-based survey that utilizes the same idea behind it as eye- and mouse-tracking methods do. The survey used shoes as an example of a consumer good and asked the respondent to state their preferences towards each shoe and then make binary choices between them. Multi-Attribute Utility Theory was used as a foundation to the study as the research aims to measure the effect of the chosen attributes (style, price and category) and the level of decision conflict. However, the main goal of the study was to test how well the chosen research method performs in this context of consumer decision process tracing. That being said, the current research should be considered as a pilot for the used research method and its results as exploratory in nature.

In the data analysis, the interest was especially on decision conflict and how that can be measured and visualized, for example, through response time and motor responses. The data collected from the survey was able to show that whenever the options differed more on the respondent's subjective preference, the response time decreased, suggesting that there was less conflict in the decision process. Similarly, a Polygon area variable that measured the respondent's scrolling movement between the options, decreased during decision tasks where the preference difference was bigger. It was also found that preference difference rating affected the proportion of responses that chose the shoe that was preferred more. The method was also found applicable for the current research setting and could definitely provide great possibilities for future research.

Keywords decision making, decision process tracing, decision conflict, multi-attribute decision making

Aalto-yliopisto
Kauppakorkeakoulu

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

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Isa Korhonen

Työn nimi Monikriteerinen kuluttajavalinta ja päätöskonflikti: Prosessiseurannan tutkimus

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Tieto- ja palvelujohtaminen

Työn ohjaaja(t) Ilkka Leppänen

Hyväksymisvuosi 2023 **Sivumäärä** 44 **Kieli** Englanti

Tiivistelmä

Päätöksentekoa on tutkittu vuosikymmenten ja jopa vuosisatojen ajan, minkä aikana myös sen tutkimusmenetelmät ovat kehittyneet. Muutama esimerkki nykyisistä suosituimmista menetelmistä ovat silmien sekä tietokoneen hiiren liikkeiden seuranta. Näihin molempiin menetelmiin liittyy kuitenkin joitain rajoituksia tai epäkäytännöllisyyksiä, jotka tekevät menetelmästä joustamattoman ja/tai kalliin. Siispä tilaa olisi uusille menetelmille. Varsinkin kosketusnäytön selailuliikkeiden jäljittäminen hiiren liikkeiden sijaan, kuten tässä tutkimuksessa testattiin, voisi olla mielenkiintoinen suunta päätöksenteon prosessien seurantaan, jotta tutkimusala pysyy mukana mobiiliin siirtyneen verkko-ostosten aikakaudella.

Tutkimus on toteutettu kvantitatiivisella menetelmällä, hyödyntäen avoimen lähdekoodin alustaa oTree:tä, jolla luotiin selainpohjainen kysely, joka käyttää taustallaan samaa ideaa kuin silmien ja hiiren liikkeiden seurantamenetelmät. Kyselyssä käytettiin esimerkkinä kuluttajahyödykkeestä kenkiä, joista vastaajaa pyydettiin ensin ilmaisemaan mieltymyksensä sekä sitten tekemään binäärisiä valintoja näiden välillä. Tutkimuksen pohjana käytettiin monikriteeristä hyötyteoriaa, sillä tutkimuksen tavoitteena oli mitata valittujen kriteerien (tyyli, hinta ja kategoria) vaikutusta sekä päätöskonfliktin suuruutta. Tutkimuksen päätavoitteena oli kuitenkin testata, kuinka hyvin valittu tutkimusmenetelmä toimii kuluttajapäätösprosessin jäljittämiseen. Nykyistä tutkimusta on kuitenkin hyvä pitää valitun tutkimusmenetelmän pilottina ja sen tuloksia luonteeltaan tutkivana.

Data-analyysi kohdistui erityisesti päätöskonfliktiin ja siihen, miten sitä voidaan mitata sekä visualisoida esimerkiksi vastausajan ja motoristen vasteiden avulla. Kyselystä saadut tulokset osoittivat, että aina kun valintatilanteen vaihtoehdot erosivat toisistaan enemmän vastaajan omien mieltymysten mittarilla, vastausaika lyheni, mikä viittaa siihen, että päätöksentekoprosessissa oli vähemmän ristiriitoja eli konfliktia. Vastaavasti monikulmioalue, jonka kuvaaja muodostui vastaajan selailuliikkeistä vaihtoehtojen välillä, pieneni päätöstehtävissä, joissa mieltymysero oli suurempi. Tutkimustulokset osoittivat myös, että mieltymyserojen suuruus vaikutti niiden vastausten määrään, joissa valittiin aluksi enemmän mieltymystä herättänyt kenkä. Menetelmän myös todettiin soveltuvan tähän tutkimusympäristöön ja vaikuttaa, että se voisi varmasti tarjota hyviä mahdollisuuksia myös tuleville tutkimuksille.

Avainsanat päätöksenteko, päätösprosessiseuranta, päätöskonflikti, monikriteerinen päätöksenteko

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and motivation

Decision making has been the topic of a great deal of research throughout decades, and even centuries, from different point of views: rational or irrational, under risk, through heuristics and biases, and based on previous references and knowledge – to name a few. Theories, models and frameworks have been created for all of them, some of which of course also contradict. The final decision itself has gained a lot of attention especially in the past, but today the research also focuses on the decision-making process, in order to better understand the *why* behind the decision. Depending on the person, the journey to the final decision can look very different from someone else's.

The decision-making process is not as straightforward as it might initially sound. It has been discovered that typically the respondent, or as its often called, decision maker, experiences some level of conflict between the options they are given (Dale et al., 2007). This decision conflict then affects the decision-making process which can again affect, for example, the response time of the decision (ibid.) Varying response times may reflect differences in the strength of preference between choice options and attributes. Therefore, in order to measure and understand how a decision has been made as accurately as possible, and especially how preferences are formed, it is important to also measure decision conflict.

A few of the most popular methods to trace the process of decision-making in laboratory settings in recent years have been the tracking of computer mouse movements (for example Dale et al., 2007) as well as eye movements (for example Milosavljevic et al., 2012) during the decision-making process. However, both of these require a lot of resources – whether it is time, money or effort – because they usually demand either specific equipment (eye-tracking), complex modeling and/or, for example, using a specialist software (mouse-tracking) (Stillman et al., 2020). Therefore, there is a need for more flexible and affordable methods, that can enable doing research experiments for a similar purpose.

New possibilities for process tracing have arisen from options like the open-source platform oTree (Chen et al., 2016), which among else, allows creating web-browser based experiments. In this research, this platform is used to develop a program that traces the respondent's decision-making process in a way that uses the same idea behind it as the mouse-tracking and eye-tracking methods do. Using this research method, the current research will explore if the information from the process tracing gathered by the program can be used to measure the strength of chosen attributes and if so, how do they affect decision conflict and the decision-making process. Therefore, the main aim of this research is to create a “pilot” for this process tracing research method, that can hopefully benefit other researchers in the future.

The thesis will focus on studying decision making in an online environment which offers a possibility to use online shopping as the decision context. Therefore, this topic is especially interesting because it can benefit both businesses and researchers. While the main focus is on shedding light to the use of the research method in consumer decision-making, it can also offer interesting insights for consumer behavior in online stores.

1.2 Structure of the thesis

The goal of is this research is to provide understanding of the use of the said experiment program in the context of consumer decision-making, how well it can measure the strength of the chosen attributes and how the possible decision conflict is visible in the decision process when using this program for process tracing. The key aspect is to test how this research method works for this type of research. The results will combine both existing literature and empirical data from a quantitative survey.

In Section 2 of the thesis, an extensive literature review to the existing literature in the field will be done. It will go through the most relevant decision-making theories for this topic as well as take a further look into the most popular research methods that have been used in decision process tracing studies.

In Section 3, the research methodology of the thesis will be discussed and the mentioned program and its use will be presented. This section will also discuss the reasoning behind the chosen research method and why it was believed to be fitting for this research.

Next, in Section 4 of the thesis, the collected empirical data and findings from it are presented. These findings will then be compared to existing literature to see where they support and contradict each other in Section 5. Lastly, in Section 6, the implications from the results for real life will be discussed, followed by discussion on limitations of the current research and recommendations for future research.

2 Literature review

The focus of this literature review will be on examining the most relevant theories and concepts regarding decision-making as well as the most common research methods for decision process tracing in similar studies. As previously mentioned, the field of decision-making is wide and multidimensional, so the aim of this literature review is to sum up the most relevant ones for this topic. Regarding the structure of the literature review, first, consumer decision making in general is discussed, followed by a presentation of the most important theory foundation used in this research, Multi-Attribute Utility Theory. Next, the most important literature and findings on decision conflict is discussed. Lastly, the most popular process tracing methods are presented.

2.1 Consumer decision making

Since the research at hand has the point of view of consumer behavior, a short introduction to consumer decision making seemed necessary. As with decision making in general, consumer decision making is widely researched field and naturally, not everything can be discussed here. However, the main aspects relevant to this research will be introduced with an emphasis on online settings.

A very typical consumer decision task involves a product or a service, accompanied with a set of alternatives, each of which are described with attributes or consequences (Bettman et al, 1998). The number of alternatives as well as attributes can vary between studies, and the chosen attributes might differ, for example, on the level of desirability or willingness to trade off, or the task might not reveal all information about the attributes to the decision maker (ibid). According to Bettman et al. (1998), the consumer decision strategies – that can also vary with the same individual – can be roughly characterized with four aspects of choice processing: “the total amount of information processed, the selectivity in information processing, the pattern of processing (whether by alternative [brand] or by attribute), and whether the strategy is compensatory or noncompensatory” (p. 189). There are several decision-making strategies that vary among these characteristics, such as weighted adding strategy, lexicographic strategy, satisficing and elimination-by-aspects (EBA) (ibid), to name a few. Bettman et al. (1998) also conclude that there are few major elements that have been discovered to affect the consumer decision making, such as the decision maker’s subjective goals, complexity of the task, context and how the task is phrased.

Existing literature ties together with the choice processing characteristics from different angles. For example, Tversky & Kahneman (1991) presented a model called reference-dependence which suggests that the decision maker’s previous reference points affect their current preference towards the options. The process of decision making can also be influenced by other factors: for example, Milosavljevic et al. (2012) looked into the level of visual saliency of the given options that can even lead to visual saliency bias, which could be seen in respect to selectivity in information processing. Dhar & Wertenbroch (2000) on the other hand, researched what kind of effect the type of the chosen consumer good had on the decision, depending on whether it was hedonic or utilitarian by nature. Gidlöf et al. (2017) discuss in their paper how consumer attention and therefore, choice, is affected by external factors (such as said visual saliency or placement of the products) and internal factors (such as brand preferences or price sensitivity) that both should be taken into consideration when analyzing decision making.

Especially in the digital focused world of Internet and online shopping, consumers are constantly bombarded with a massive load of information, and making good decisions in that environment can become difficult (Fasolo et al., 2007). This information overload can even result into consumers’ unwillingness to actually make a decision and to dissatisfaction with the decisions they eventually made (ibid). Fasolo et al. (2007) found that especially the number of product attributes present in the decision-making situation can negatively affect the consumer and create bigger decision conflict, in comparison to having more options with less attributes. While physical stores have at least some limitations to what attributes they can present with the product (for example price or materials), online stores can theoretically present as much information as they want. Typically, when the decision difficulty increases with number of attributes, consumers turn to heuristics, such as Take The Best -heuristic, to make the decision easier for themselves (ibid). Niza Braga & Jacinto (2022) also found that online shopping environment can increase the use of heuristics, due to the consumer’s expectation of online shopping to be low-effort which can prime the use of resource-saving goals.

2.2 Multi-attribute utility theory

The main decision-making theory utilized in this thesis is Multi-Attribute Utility Theory (MAUT). It lays its' foundation on Expected utility theory but elevates it to facilitate several attributes. Multi-Attribute Utility Theory is a Multiple Criteria Decision Analysis (MCDA) approach which is essentially about aiming to allocate a utility value to every action by the decision maker (Dyer, 2016). The assigned utility value is a number that represents the preferability of each of those actions and is often "the marginal utilities that each criterion assigns to the considered action" (ibid, pp. xxvi).

Although neither Expected utility theory nor MAUT can be thanked for by just one researcher as these have been developing especially throughout the last 100 years, for example the works of Peter Fishburn with the former and of Ralph Keeney & Howard Raiffa with the latter theory, have become very popular among researchers (Dyer, 2016). The studies using utility theories can be divided into two kinds: the ones where the alternatives are portrayed by just a single attribute and the ones by multiple attributes (ibid), out of which the second option is used in this research. There also exists a vast amount of previous research using multiple attributes. For example, Fasolo et al. (2007) conducted a simulation study with a total of nine attributes in their research about the optimal level of attributes. Cohen et al. (2017) on the other hand, implemented a part of their research on multi-attribute, multi-alternative models of choice with four attributes.

Additionally, the multi-attribute models can be investigated either from the viewpoint of certainty or risk (Dyer, 2016). For example, Koop & Johnson (2011) used a common type of decision task under risk, The Iowa Gambling Task, to study response dynamics. A generally adopted distinction between the two viewpoints created by Keeney and Raiffa in 1976, uses the term "value function" for preference functions under certainty and "utility function" for preference functions under risk, which also Dyer (2016) uses, that also refers to MAUT as multi-attribute preference theory to cover more multi-attribute models of choice. As they state, "preference theory studies the fundamental aspects of individual choice behavior, such as how to identify and quantify an individual's preferences over a set of alternatives, and how to construct appropriate preference representation functions for decision making" (ibid, p.287). A significant element of preference theory is also that it builds on strict preference axioms that both characterizes the decision maker's choice behavior as well as creates essential foundation for the quantitative analysis of preference (ibid).

The empirical research for this thesis will also include more than one attribute and therefore, MAUT works as a foundation to it. These multi-attribute, multi-alternative situations are very common in real life; most often when making a decision, several aspects need to be taken into consideration, which makes this an interesting and practical point of view. This research will also focus on the preference presentation of certain options where the decision maker knows the necessary information for making the decision, in comparison to risky options which would include lotteries and gambles where some level of probability for wins and losses is presented or is unknown. When discussing preferences under certainty, we are especially interested in the strength of the preference (Dyer, 2016) in order to dig deeper into the why behind the final decision as discussed above.

2.3 Decision conflict

In life, some decisions are easy and some are not, but usually there is some level pondering between the available options, and even though people might end up making the same decision or for example categorization in the end, their route to that decision might be very different. Especially when there are multiple different aspects, or attributes, to consider, the level of decision conflict between the options might increase. According to Bettman et al. (1998), the difficulty of a decision task rises, for example, with the increase of either alternatives or attributes (or both), with higher uncertainty regarding the attribute values given that there are more attributes that the decision maker finds hard to trade off, or if the given alternatives share a smaller number of attributes.

Stillman et al. (2020) defines decision conflict as “the lack of strength and consistency of the signals toward one option over the other” (p. 31738), where by signals they mean internal indications or evidence towards the dominant option which can be either high or low depending on the level of conflict. A little more simplistic explanation by Stillman et al. (2018) defined it as “the relative amount of conflict when deciding between two possible choices. For instance, self-control decisions are difficult owing to the decisional conflict between short-term and long-term gratification” (p. 532). The matter of decision conflict has been researched in the past literature from many angles and the similar matter has been discussed with different terms. For example, Dale, et al. (2007) discuss complexity in their research using mouse movement trajectories that indicate the route of the mouse cursor towards the chosen option and how straight the line was i.e. how much (or little) decision conflict was involved in the process. Cheng & González-Vallejo (2018) on the other hand talk about decision difficulty, which they divided into conflict and wavering, of which level of strength was affected by contextual changes. Philiastides & Ratcliff (2013) use in their research about the influence of branding on decision making the diffusion model that entails internal components of processing which include for example nondecision processes.

Chatterjee & Heath (1996) also discuss in their paper about decision difficulty which is affected by the amount and type of decision conflict. In their experiments, Chatterjee & Heath showcased the decision maker two options that seemed fairly attractive (approach-approach conflict) or two that seemed fairly unattractive (avoidance-avoidance conflict), while including attribute trade-offs within the presented options (embedded approach-avoidance conflict) (ibid). They also nicely discuss how many decision difficulty studies revolve around computational difficulty that arise due to the processing of multiple attributes and when the computational complexity or overall information load becomes overwhelming, decision makers turn to for example, heuristics, in an effort to make the decision making easier (ibid), as discussed above as well. Tversky & Kahneman (1974) thoroughly present a set of heuristics, and biases, in their paper that show how these can be beneficial by simplifying decision making, but can also result into some potential pitfalls and errors in thinking. Within the topic of information overload and decision conflict, Fasolo et al. (2007) talk about ”tyranny of choice” and argue that having fewer attributes can in fact make the decision easier while still being able to make good enough decisions in everyday life.

As said, conflict or difficulty in decision making has been studied with many names depending on the aim of the study, but in this research the term decision conflict will be used from now on for clarity. However, the results from the empirical, quantitative research will be reflected even with previous literature that does not use the exact same term.

2.4 Current research methods in process tracing

This chapter of the literature review will take a look into currently popular research methods within the field of process tracing through previous literature. There exist several process tracing methods but few that come across often are mouse-tracking and eye-tracking methods. These process tracing methods allow “the observation of cognitive dynamics in action and yield[s] temporal information about the development of a response" (March & Gaertner, 2021; p. 2439). The said methods can, for example, look into "bottom-up" factors and "top-down" factors (Hehman et al., 2015; Chocarro et al., 2022). The bottom-up factors arise from the task itself, such as the visual characteristics of the options, while top-down factors are cognitive in nature and come from the respondent's side, such as attention, motivation or knowledge (Hehman et al., 2015), similarly to what Gidlöf et al. (2017) referred to as external and internal factors.

2.4.1 Eye-tracking

The human gaze has held a large interest in past research due to its fundamental part in the way we communicate as well as for our overall cognition (Bulling & Wedel, 2019). It has been researched from various points of views, for example, within the fields of psychology, sociology and consumer behavior (ibid), the last one also being the focus of this research. According to Bulling & Wedel (2019), the ways of assessing the human gaze, attention or movement patterns of the eye during a period of time, "use either sensors placed in the environment (so-called stationary eye-tracking) or worn on the head (so-called mobile eye-tracking)" (p. 27). Either way, very specific equipment is required. However, improvements in the needed technology with mobile eye-tracking equipment have enabled the eye-trackers to be become lightweight embedded systems that allow recordings in everyday settings (ibid). For example, Gidlöf et al. (2017) conducted a research on how consumers' visual attention and the final choice are influenced by the subjects' preferences and the elements of a supermarket shelf by giving mobile eye-trackers to a group of consumers visiting a real supermarket and tracing their attention during grocery shopping. For this study, the researchers used eye-tracking glasses that the subject wore during the shopping trip (ibid). The stationary eye-tracking technology has also seen big developments throughout the years, allowing the use of only one camera, even in a hand-held device Bulling & Wedel (2019). These developments enable reliable tracing of eye movements that is continuous and robust in daily situations instead of just laboratory settings (ibid). All in all, the idea behind the methodology is to follow the movement of the eyes to better understand how the subject comes to their decision and where are their eyes directed during the decision-making process.

The first eye-tracking devices were developed already at the end of 1800s, but the first editions of course still contained quite large limitations (Bulling & Wedel, 2019). According to Bulling & Wedel (2019), today eye-trackers usually record eyes' movements in one of the following three ways: “a) video-based infrared pupil-corneal reflection (PCR), b) measurement of the cornea-retinal standing potential between the front and back of the human eye (Electrooculography, EOG), and c) video-based eye-tracking using head-mounted or stationary visible light video cameras (video-oculography)” (p.29). For example, Krucien et al. (2017) used the option c – video-based eye-tracking system as the stationary version in their research about visual attention in multi-attribute choices. The frequency of the recordings is very high with the eye-tracker they used as it registers the eye movements at 1000Hz rate, which means 1 recording every millisecond (ibid). However, this rate can of course vary with different equipment. For example, Wang et al. (2014), used equipment that had a recording rate of 500 Hz, while Chocarro & Villanueva (2022) utilized an eye-tracker with an average rate of 30 Hz.

The attention of the decision maker has in fact been an important topic within the eye-tracking field. In their research, Orquin & Loose (2013) found that attention process seems to pose an active part during the constructions of a final decision. Also, Zuschke (2020) stated that the main purpose of eye-tracking within decision making is to measure visual attention. The data collected from eye-tracking about attention has been further used to research for example dwell time, meaning a longer gaze at a certain point. For example, Gidlöf et al. (2017) counted gazes longer than 100ms as dwell time. Shorter gazes i.e. quick eye movements, are called saccades and according to the findings of Gidlöf et al. (ibid), the longer the dwell time, the more probable is buying the product. Krucien et al. (2017) also looked into dwell time, although they used the term fixations, which they considered to be gazes longer than 50ms, while Chocarro & Villanueva (2022) defined fixation as more than 200ms. No one definition for the length seem to exists as it varies between studies.

In addition, it seems to be quite typical to also collect data in some other method than just eye-tracking. For example, Gidlöf et al. (2017) also conducted a questionnaire on the same subjects who participated to the eye-tracking section of the study, in order to get some insights on the subjects’ former familiarity with the supermarket the research was done at and how different attributes of the products had impacted their purchase decision. A similar combination was also used by, for example, Wang et al. (2014) and Chocarro & Villanueva (2022), while Boardman et al. (2022) used qualitative semi-structured interviews to get further interpretations to enrich their eye-tracking data. Milosavljevic et al. (2012) on the contrary, decided to repeat their eye-tracking experiment with hand movements to validate their results, where instead of looking at the option the subject wanted to choose, they were asked to press a keyboard's left and right arrow keys.

Some shortcomings are however still apparent with the modern-day equipment. According to Gidlöf et al. (2017) even considerable data loss and inaccurate data from the recordings occurs often with mobile eye-trackers as the eye-tracker can move around enough to become dislocated or lose its calibration, or uneven lighting conditions might negatively affect the recording. Krucien et al. (2017) for example, had decided to use head rest for the participants in order to avoid too big of head movements that the camera could no longer record the eye movement correctly, although this would not of course be possible outside the laboratory setting. Also, there is of course always the inconvenience of needing specific equipment for the experiment, even when using lightweight mobile eye-trackers.

2.4.2 Computer mouse movement tracking

Computer mouse movement tracking is also a widely used method for decision tracing that, as the name suggests, tracks the movement of a computer mouse to better understand the decision-making process. It allows the researcher to examine in detail the nature of the decision conflict, as well as how the decision evolves and the conflict is resolved over time (Stillman et al. 2018). According to Stillerman & Freeman (2019), usually mouse-tracking experiments showcase the subject a screen with a start button at the bottom-center of the said screen, that by clicking reveal the stimulus in question, after which typically two options appear in the top left and right corners. However, the form of the experiment can of course vary – for example Koop & Johnson (2011) had a total of four options in each corner of the screen with the start button appearing in the middle of the screen. When the cursor of the mouse is moved on screen, x- and y-coordinates are recorded which can then be visualized as mouse-trajectories over time course (Hehman et al., 2015).

The core of today's mouse-tracking research lies in dynamic motor responses. These response dynamics can capture "the continuous, online processing of information as it is revealed in the subject's motor response" (Koop & Johnson, 2011: p. 751). The research by Spivey & Dale (2006) has indicated that humans' cognitive processing can be discovered in the dynamic motor responses of mouse-tracking paths that are able to measure the preference evolution that eventually leads to a choice. The general paradigm of response dynamics involves the recording of the mouse cursor's position (x- and ypositions as mentioned above) on its way to the selection of one of the options (Koop & Johnson, 2011). The idea is that this method would be able to also track momentary preferences towards other options before deciding on the final choice and therefore offer a continuous insight to the decision-making process (ibid). The experiment also typically has time limits for starting the choice task and for making the actual choice in order to create pressure for the subject to start making their response before they have actually made the choice in their mind, which is expected to show the decision process better in the computer mouse movements (Stillerman & Freeman, 2019). In their research about motor responses during categorizing atypical exemplars, Dale et al. (2007) showed the given options to the respondent for 2 seconds before they had to make their decision. Lee et al. (2021) on the other hand, showed the options for 3 seconds in their research about fashion advertisements.

One of the most used measurement within mouse-tracking studies that also lays its' foundations on the idea of response dynamics, is Area Under the Curve (AUC) which counts the area between the mouse-trajectory from the respondent and an idealistic linear line from the start point to the chosen option (Koop & Johnson, 2011). The aim is to assess the subject's attraction towards the other, non-selected option – so in other words, the decision conflict – over the response time (ibid). Stillman et al. (2020) found in their research that mouse-tracking was able to precisely quantify conflict, at least in risky choices (gambles). Another common measurement or analysis model is the drift diffusion model (DDM) which for example Stillman et al. (ibid) used in their mouse-tracking research about predicting and understanding risk preferences. This model has the assumption that over the period of response time, the respondent gathers and compares their internal signals towards the options until one of the options reaches a certain, predetermined threshold (ibid). As with eye-tracking, depending on the equipment – here a software – the frequency of the x-, y-coordinate recordings varied between studies but, for example, Dale et al. (2007) had an average rate of 42Hz in their research while Koop & Johnson (2011) had a recording rate of a 100 Hz. These mouse movements can, for example, be predictive of the subject's risk preferences in a decision task even if the final choices wouldn't be (Stillman et al., 2020).

Mouse-tracking seem to provide advantages in comparison to other process tracing methods like eye-tracking, such as providing a readily interpretable, dynamic evaluation of choice, offering an inexpensive and easily accessible method for researchers, and allowing to scale the research of conflict outside of laboratory setting (Stillman et al., 2018; Stillman et al., 2020). However, it should be taken into account during the design process of the experiment that subjects might develop strategies where they simply travel the mouse cursor in middle of the screen rather than moving it towards the options to “buy” thinking time (Maldonado et al., 2019).

2.5 Research questions

Given the findings and the state of understanding of process tracing as well as decision conflict from previous literature, the research questions for this research were defined as follows:

1. Can the information gathered from this decision process tracing method be used to measure decision conflict and define the strength of the chosen attributes?
2. How does the level of decision conflict impact the respondent’s decision-making process, such as via response time?

3 Methodology

The empirical part for this thesis was done as a quantitative research in order to add on new findings to the existing literature. The chosen quantitative method was an online survey of which results were further investigated using programming languages R and Python in Jupyter Notebook and Google Sheets. The survey was done using an open-source, online software oTree (Chen et al., 2016) that allows for implementing interactive experiments, such as web-based surveys. This section of the thesis will first present the background for choosing this research method, followed by a discussion on the survey design. Lastly, the goals of the research are presented.

3.1 Background for choosing the research method

The initial selection of a quantitative research method arose from the possibility of gathering a larger set of respondents that would then of course offer a bigger set of data to make conclusions from. Decision making has already been researched in various ways, by numerous different researchers and the aspects looked into in this thesis are not the type that one could get for example from interviews. Hence, gathering a bigger data set with a quantitative survey was a more suitable option here.

As discussed in the literature review, decision process tracing has heavily relied on mouse-tracking and eye-tracking methods in the past. However, both of these require, for example, specific equipment or software, in order to trace the decision-making process which makes both of them somewhat inflexible and even expensive. Therefore, other ways are required. oTree by Chen et al. (2016) builds on the highly popular platform z-Tree by Urs Fischbacher (2007) that can be used with desktop computers which have a Windows operating system. However, as it already sounds, this creates quite strict limitations to the use of z-Tree and hence, oTree was created in way that allows for any device, operating system and browser.

As importantly noted by Chen et al. (2016), consumers' usage has greatly shifted away from desktop computers to hand-held devices such as mobile phones and tablets which has also resulted into the revolution of high-quality graphics and interfaces which is why they wanted to develop a software that is deployable in all of these devices and interfaces. oTree platform can be used for many kinds of experiments from laboratory to field to online studies and in this research, the last one will be utilized. This offers a possibility to develop a program that can be used in any device that is connected to internet and on any browser, as an interactive survey that doesn't require the respondent to install or log in to anything. oTree's user interface uses HTML5 as its foundation along with Java Script and Cascading Style Sheets (CSS), and the programming language used for the programming itself is Python (ibid). oTree is also built on the Django web application framework, meaning that the experiments done on oTree are web applications (ibid).

oTree was chosen as the platform for creating the survey since it seemed to offer all necessary features that maybe more traditional, purely survey platforms did not have. The used program and the code itself were developed by utilizing and combining existing pieces of code while modifying it to the current use case. By developing this program on oTree, the aim is to shed some light to using this method as an alternative to eye- and mouse-tracking for decision process tracing.

The idea behind eye- and mouse-tracking methods is that it is able to gather information on the respondent's decision-making process. For example, information on where the decision maker looks, how long of a gaze they have on the options, what kind of movements are done using their computer mouse and how much time the decision-making takes, can be collected and analyzed. The same idea of following the decision-making process is also utilized in the research method used for this thesis. However, in this case, it will be done with a survey using oTree that collects data of the response process. In general, the main advantages of using an online survey as a research method are the low costs, possibility to gather large dataset fast – from certain target group if needed – and high quality of data (Nunan et al., 2020). Also, with surveys, no interviewer bias is present (ibid). However, some disadvantages can be, for example, respondents limited access to Internet and possible technical difficulties (ibid).

3.2 Survey design

The setting of the survey was inspired by online store environment and decisions made in online situations. The focus of this research and the survey was on better understanding consumer's decision making and therefore, the survey utilized a common consumer good, shoes. It was chosen as an example product since it could be said fairly confidently that it is a product that is well-known to all respondents, regardless of demographic background. All the presented pairs of shoes were from one brand, the sportswear giant Nike, for the purpose of eliminating the respondents' existing preferences towards different brands. The respondents might of course have opinions and preferences towards Nike, but this way the preference remains the same throughout the different options. The presented pictures of the said shoes were screenshots from the selection of Nike’s online store nike.com and were taken on 8.5.2023. The screenshots were taken from the brand’s “running” and “lifestyle” categories, and included shoes from their men’s, women’s and unisex collections. In fact, many of the shoe models could be found from both men’s and women’s selections. For the survey, 20 pairs of “lifestyle” shoes and 20 pairs of “running” shoes were selected. The attributes that were chosen for this research were categorization of the shoe to lifestyle versus running usage, the look of the shoes (style) and the price. These attributes were chosen because they are easily understandable and clear for the respondent and therefore, good for comparing.

The purpose of the survey was to look into the research problem described above. When investigating the past literature, the same type of survey using oTree had not been used for studies on consumer’s decision-making behavior and how well the strength of different attributes and decision conflict can be measured with this method with consumer goods. Therefore, it was interesting to explore if this method suits consumer decision process tracing in a way that allows for better understanding of the respondent’s decision conflict between the alternatives and chosen attributes.

The programmed survey consisted of two parts: the “base” questions and decision-making situations which were again divided into three sets. The first part included initial questions in order to gather information on the decision maker’s, or respondent’s, preferences between the shoes. The respondent was given a picture of a shoe and asked to indicate how likely would they wear that particular shoe on a 7-point Likert scale (-3 extremely unlikely, 3 extremely likely). In this part, the respondent was asked to state their preference only based on the style / appearance of the shoe. This preference score was given to all of the 40 pairs of the shoes included in the survey.

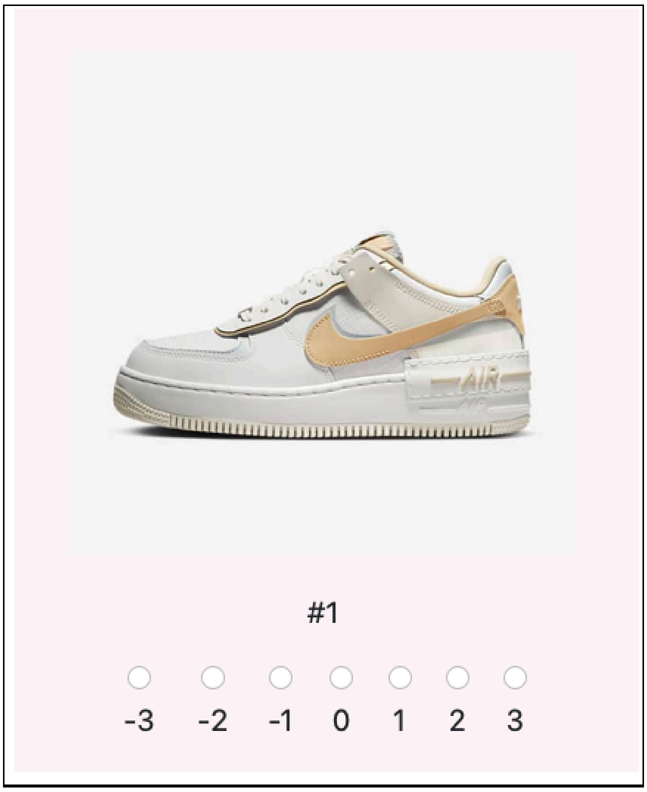

Figure 1. Example of preference statement question (instructions not visible here).

The second part of the survey included several decision-making situations, where the respondent had to choose between two options of shoes. The two alternatives for each decision task were randomized with the program in oTree for each of the three sets. By randomizing the pairs, it was made possible to have some pairs that were very similar and some that were very different in the participant's recorded preference towards them. The survey included a total of 40 decision tasks. This number of rounds were chosen in an aim to gather enough data from each respondent, but to keep the survey short enough as the respondents were not compensated for their responses. With the 40 rounds, the total time to finish the survey was about 10 minutes. See Figures 2-4 for illustration of the decision tasks for each set. In all of the 40 rounds, the respondent was asked to swipe the empty, grey boxes shown to them in order to reveal the two options and their possible additional information. The options were not fully visible at same time, meaning that the respondent had to scroll back and forth between the options. The final decision was made by scrolling to the option they preferred and then clicking OK.

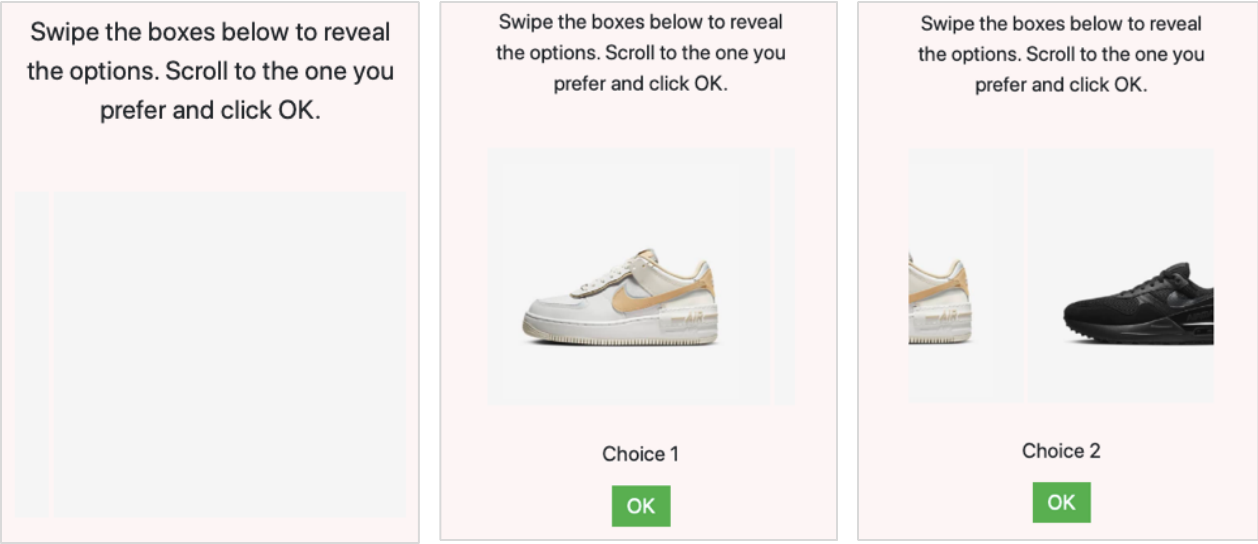

Figure 2: Example of a choice task with style attribute

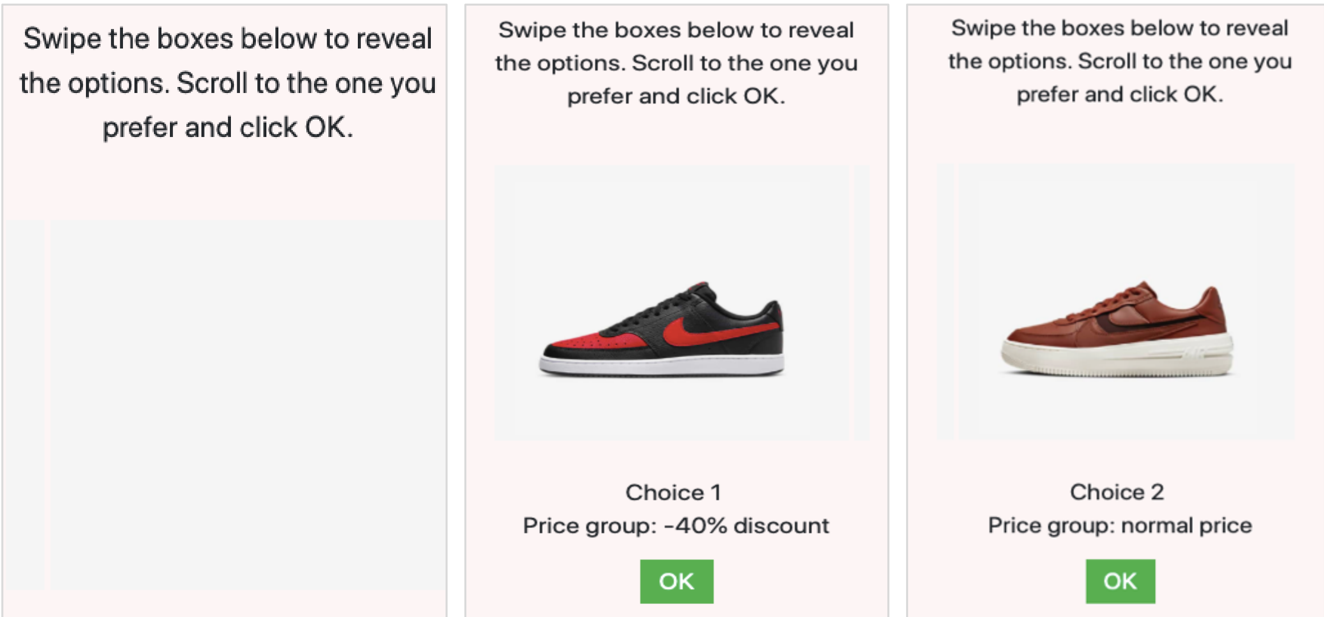

Figure 3. Example of a choice task with price attribute

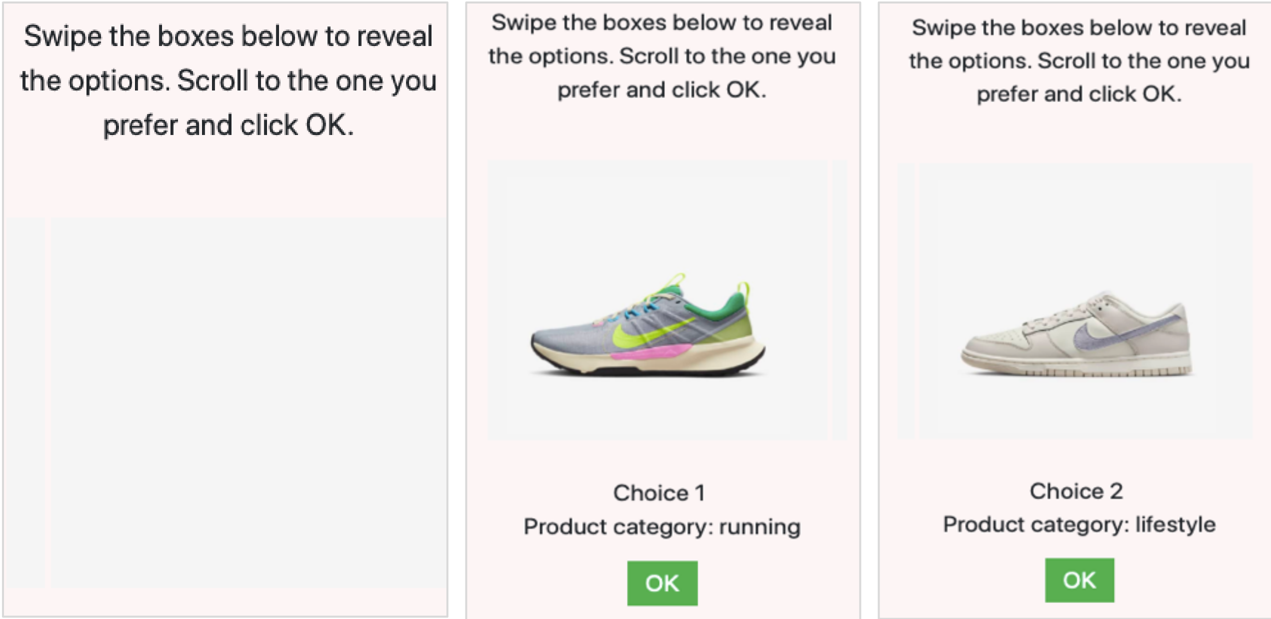

Figure 4. Example of a choice task with category attribute

In the first set (see Figure 2), the respondent was asked to make their decision again based on only the style / appearance of the shoes, for a total of 14 rounds of decisions. The second set (see Figure 3) gave the respondent also information on the price level of the product, in addition to the picture of the shoes. The price level had two categories: "-40% discount" and "normal price". The prices were chosen to be shown as categories instead of the real prices since the respondents might have a very different opinion on what is expensive and what is not, so presenting the price might have ended up affecting the results in a manner that was not measured. However, price as an attribute tends to possess a generalization that lower price is better in all cases, if no other attribute is considered. These particular price categories were chosen because it was assumed that these would present a more realistic feeling towards the price point rather than just high/low. The pairing of a price level and certain shoe was also randomized with the survey program and was not based on real prices in Nike's online store. The respondent was asked to make decisions based on the picture (style) and price level for 13 rounds. The third set of decisions (see Figure 4) asked the respondent to make the decisions based on the picture and the product category (lifestyle / running) where the shoe belongs to in Nike's online store. The respondent was also advised that they are looking to buy shoes for everyday use. This set also included 13 rounds of decision tasks.

3.3 Goals of the research

The programmed technique behind these decision tasks was aimed to trace the decision-making process. The program for example tracked how long it took for the respondent to make the decision from swiping the empty boxes to clicking OK for the chosen option, as well as, the coordinates of the scrolling between the options. It also saved the information on which shoe pairing each of the participants were given and their corresponding preferences from the first section of the survey. Here, the main goal was to see how the amount of decision conflict between the options affected the decision-making process. For example, the respondent could have been shown two shoes that they had reported to highly prefer. In addition, as said, this second section of the survey was divided into three sets in order to test how well this research method worked for measuring the strength of the chosen attributes, style, price level and product category.

The model of the survey was designed based on previous literature, such as Philiastides & Ratcliff (2013), Milosavljevic et al. (2012) and Enax et al. (2016). However, since the topic of each of these was a little different than of this research, the survey questions and details were of course modified for its' needs. For example, Philiastides & Ratcliff (2013) studied how branding affects preference-based decision making. In their experiment, they first asked the participants to rate the clothing items shown to them on a 7-point Likert scale (ibid). For the purpose of their study, Philiastides & Ratcliff (ibid) also asked the participant to rank 24 brand logos based on their preferences towards the said brands. The main task in the study was to make binary choices between less-preferred and more-preferred brands. To test the strength of the "brand" -attribute, they presented pairings of pieces of clothing with and without the brand logo (ibid), similar to how in this research the participants were given binary choice tasks with just the photo, with price information and with category information.

4 Data & Findings

In this section, the quantitative data from survey and the findings from it will be presented. First, some general information and statistics will be discussed, followed by more thorough insights found from the data. The statistical testing of these findings will also be presented. As mentioned before, the current research should be considered as a "pilot" for the research method and the collected data related to it should also be considered as exploratory in nature.

The data for the survey was collected using a nonprobability sampling method called convenience sampling where the sample is selected based on the convenience of finding respondents (Nunan, et al., 2020). The data was collected on 9.6.-12.6.2023 by sharing the link on the researcher's social media account and on other communication platforms. There was a total of 79 respondents who successfully completed the survey, meaning that N=79. Since each respondent did 40 rounds of decision tasks, this meant data from a total of 3160 tasks. This sample size seemed reasonable in comparison to previous studies using eye-tracking (e.g. Wang et al., 2014; Boardman et al., 2022; Chocarro et al., 2022) and mouse-tracking (Dale et al., 2007; Koop & Johnson, 2011; Cohen et al., 2017) as this sample included even more respondents. For example, Nielsen Norman Group (2021) suggests that a sample size of about 40 participants is usually valid and reliable enough for a quantitative study. It was brought to the researcher's attention that a few potential respondents had experienced some technical issues during the survey and weren't therefore able to finish it. Hence, only the responses that were complete are included in the analyzed dataset. For future use of this research method, these possible technical issues should be kept in mind and eliminated where possible.

The handling and analysis of the data was conducted using few different methods – coding languages R and Python with Jupyter Notebook and Google Sheets – depending on the need. Before the analysis could have been started, a clean-up and re-arrangement to the dataset was necessary in order for it to be ready for analysis. The process began by deleting all columns of data that weren't needed for this analysis that the oTree survey produced, after which both the price attribute labels (normal price, -40% discount) and category labels (lifestyle, running) were encoded to numbers i.e. dummy variables (1 and 0 in both respectively). One main goal of the analysis was to see how the respondents' preferences towards each shoe affected the decision making in the binary choice tasks. Therefore, a new variable was created that measured the difference between the preferences of the given options in each task. In this research, it was called preference difference, but for example Philiastides & Ratcliff (2013) talked about difference ratings which they referred to as “task-difficulty level” to signal how a bigger difference between the respondent’s stated preference towards the product options was assumed to be an easier choice, while small difference implicating the opposite.

After this, several subsets of the data were created based on different criteria to help the analysis process. First, two datasets were extracted from the entire data: one with the choices where the preference difference between the presented options in each choice task was low and one where the difference was high. For this research, it was determined that the “low preference difference” -category (presumably more difficult tasks) included choices where the difference was either 0 (no difference) or 1, while the “high preference difference” -category (presumably less difficult tasks) included choices where the difference was 3 or more. For this purpose, the difference ratings were utilized as absolute numbers. Next, three more subsets were created from the data: one that included the choice tasks that only had the style attribute i.e. only presented the pictures of the shoes, one that included the choice tasks which had the price attribute present in addition to the pictures, and one that included the choice tasks that had the category attribute present in addition to the pictures. All these subsets were then further divided into subsets based on the preference difference, as was done to the entire data, that could be used for further analysis.

4.1 Behavioral data analysis

4.1.1 Response time

To start with, the response time (RT) of the responses was analyzed for different scenarios, based on the chosen attributes (style, price, category) and the subjective preferences stated in the first section of the survey. Response time has also been called, for example, reaction time and decision time in the existing literature, but all correspond to the same measurement. Response time consist of the processing time (conflict) where the alternatives are evaluated and the final decision making. The analysis of response time was included to the analysis section to see if the level of decision conflict could be seen in the response time. Because the mean response times can be sensitive to outliers (Enax, et al., 2016), this data also had to be cleaned a bit as those outlier responses were most likely not done in full focus and therefore, affected by non-attention and distraction that led to problematic data that distorted the mean response times. To do so, the responses with longer than 30 seconds where initially excluded to leave out the biggest outliers. Then, for the remaining rows (3113; 98.5% of the entire data), mean response time and standard deviation were checked, which were 6.22s and 4.15 respectively. The maximum limit for response times was then defined by adding two standard deviations to the mean that equaled to 14.51s, which felt reasonable to be rounded up to 15 seconds. A similar process was also done by Enax et al. (2016), although they measured the maximum cutoff point for each respondent separately. After excluding all responses that had a response time longer than 15 seconds, a total of 2966 decision tasks remained in the dataset, meaning that 194 responses were excluded, which corresponds to 6.1% of all the responses.

All of the average response times are summarized in Table 1 below. For the entire dataset that now included the 2966 binary choices from the 79 respondents, the average response time was 5.53 seconds. In addition to measuring the average response times, t-tests against the entire data set was performed to see if the change in response time for each attribute and for low and high preference difference tasks was statistically significant, where $p < .05$ was considered significant. In addition, t-tests were performed within each attribute subset for responses with low preference difference and high preference difference. The exact p-values are also summarized in Table 1. Due to response time distributions being highly skewed, the response times were log-transformed to be closer to normal distribution, as did Enax et al. (2016) in their research. The low difference subset within the entire dataset included 1356 choice tasks and the average response time from these was 6.01 seconds ($p < 0.01$). The high difference category in turn had 1022 choice tasks and average response time was 4.94 seconds ($p < 0.01$). Both of these changes were therefore statistically significant.

Next, the mean response times were measured for each attribute subset and the low and high preference difference scenarios within them. For the style attribute only, the dataset included 1023 choice tasks and the mean response time was 5.5 seconds ($p > 0.05$). For the low preference difference tasks within the subset, mean response time was 6.15 seconds ($p < 0.01$) and for high preference difference tasks 4.86 seconds ($p < 0.01$). Hence, the change for entire style-subset's average response time was not significant, but the low and high difference scenarios within the subset were. For style + price choice tasks, the dataset included 959 tasks and the mean response time was 5.81 seconds ($p < 0.01$). Within this subset, for low preference difference tasks the mean response time was 6.31 seconds ($p < 0.01$) and for high preference difference tasks 5.21 seconds ($p < 0.01$). Here the entire price subset as well as the low difference tasks and high difference tasks within that subset presented a statistically significant change. Lastly, for the style + category choice tasks, the dataset included 984 tasks and the mean response time was 5.29 seconds ($p < 0.01$). For low preference difference tasks within the subset, the mean response time was 5.59 seconds ($p < 0.05$) and for high preference difference tasks 4.76 seconds ($p < 0.01$), meaning that all of them were statistically significant, but the low difference tasks within the subset were a little less significant.

As can be seen, the mean response time within each section was always lower when the preference difference was high than when the it was low. The lower response time when the preference difference was higher could indicate that the respondent experienced less decision conflict between the options.

Table 1: Summary of the response times for different scenarios and p-values againts the entire data and corresponding subsets

Scenario	N	Average response time	p-value
Entire data	2966	5.533	-
Entire data: Low preference difference	1356	6.008	2.632e-09**
Entire data: High preference difference	1022	4.938	3.001e-12**
Style attribute (baseline)	1023	5.501	0.2836
Style attribute: Low preference difference	1023	6.146	0.0000151**
Style attribute: High preference difference	1023	4.865	0.00002016**
Style + price attribute	959	5.813	0.00007225**
Style + price attribute: Low preference difference	959	6.315	0.0003946**
Style + price attribute: High preference difference	959	5.213	0.00009823**
Style + category attribute	984	5.294	0.009026**
Style + category attribute: Low preference difference	984	5.594	0.01048*
Style + category attribute: High preference difference	984	4.756	0.0001017**

4.1.2 Polygon area

In addition to response time, the Polygon area (see Figure 5 for an illustration) where the decision maker's conflict is visually and numerally expressed and calculated as path integral, was analyzed. When answering the survey, the program collected both x- and y-coordinates of the decision-making process, i.e. the scrolling between choice 1 and choice 2 was recorded as part of the process tracing. The grey boxes visible in Figures 2-4 are HTML-elements and tracing of their movements is based on an onscroll event that asks a JavaScript function to register an x-coordinate whenever the box is moved to left and right. The x-coordinates are expressed as pixels relative to the midpoint of zero of the mobile device. While each x-position is recorded, a timestamp is also recorded that creates a y-position. Hence, the Polygon area is a tempo-spatial measurement that takes into consideration both the amount of scrolling movement as well as its speed. This information was recorded for each respondent and for each decision task they made. The frequency of x- and y-position recordings can vary based on the device and browser being used, but in this data set, the average was 17 recordings per second, which translates to 17 Hz. From the recorded x- and y-coordinates, one row was excluded from the data and altered to N/A due to having a value of 0, which indicates that the respondent did not view both options properly before making the decision.

An example of what the visualization of polygon area looks like can be seen from Figure 5. The line begins from the bottom the graph, from a randomized starting point and ends at the top of the graph either to the left or right side, depending on if the respondent chose option 1 or 2. From the graphs, it is easy to see the scrolling process the particular respondent had during these rounds, which all were style attribute tasks in these examples. For example, in round 6 (top right graph), the respondent spent a great deal of time scrolling between options while with round 10 (bottom right graph), the decision was quite straightforward. Here, the respondent has chosen option 1 (left side) in all of these rounds except for the bottom left graph.

The idea with the Polygon area measurement was to analyze the decision makers' conflict from the angle of motor responses. In addition to the mere x, y -coordinates, the area inside the path was calculated in order to make further analyses. By calculating the area within the path that is connected with a straight line between the starting and ending point to create a closed plane figure, it was aimed to measure the decision-making process in a similar manner as Area Under the Curve (AUC) measurement, which has been used with, for example, mouse-tracking methods (Hehman et al., 2015; Stillman et al., 2020). While response time is also a good metric to measure, some existing literature suggests that response time might be more influenced by motor latencies than for example AUC (Stillman et al., 2020) and hence, it was interesting to compare the findings from response time as well as from Polygon area.

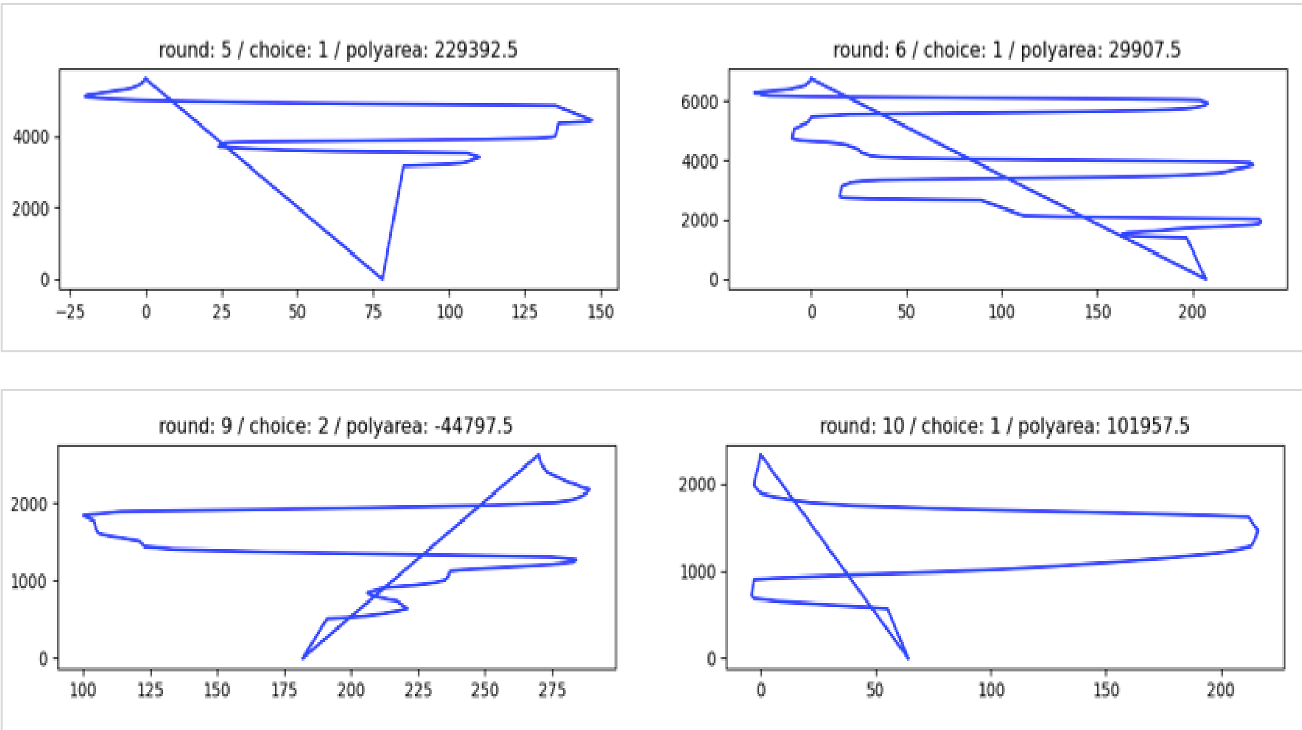

Figure 5. Examples of Polygon area visualization as path integral

To gain a better understanding of the Polygon area variable, a similar analysis was done as previously for response time. The mean values for each scenario were calculated and the statistical significance of the change against the entire data as well as low and high preference difference tasks against their subset were tested with a t-test. These measurements are summarized in Table 2. Since the Polygon area measurement could initially be either negative or positive depending on the direction of the integral function used to calculate the area – which does not describe the size itself – the measurements were transformed to absolute numbers for better comparability. With the Polygon area variable, the actual size of it is not of interest and even that understandable measurement on its own – instead, the relationship or comparison of the sizes should be considered. As mentioned, a bigger Polygon area means more scrolling between the options which again indicates a bigger decision conflict.

Similar to the response time comparisons, the Polygon areas from the decisions with low preference difference were bigger than the ones with high preference difference in each of the scenarios. For the entire data the mean Polygon area was 303144.10, while for low difference tasks it was 331324.84 ($p < 0.01$) and for high difference tasks 265502.34 ($p < 0.01$). Based on the p-values, both scenarios were statistically significant. For style attribute subset, the mean Polygon area was 300996.97 ($p > 0.05$), while for low difference tasks within this subset, it was 343584.96 ($p < 0.01$) and for high difference tasks 255952.79 ($p < 0.01$). While the change was not statistically significant for the entire style attribute, it was for the low and high difference tasks within the subset, which was similar to response time for the same scenarios. For the price attribute, the mean Polygon area was 324271.0 ($p < 0.05$), while for low difference tasks within this subset, it was 356355.07 ($p < 0.05$) and for high difference tasks 290022.63 ($p < 0.05$). Here, all of the scenarios were statistically significant. Lastly, for the category attribute, the mean Polygon area was 284767.52 ($p < 0.05$). Within this subset, the Polygon area for low difference tasks was 296520.31 ($p > 0.05$) and for high difference tasks 252827.24 ($p < 0.05$). The low difference tasks within the subset were not statistically significant, but the other two were.

Table 2: Summary of the Polygon areas for different scenarios and p-values againts the entire data and corresponding subsets

Scenario	N	Polygon area	p-value
Entire data	2966	303144.10	-
Entire data: Low preference difference	1356	331324.84	0.0008258**
Entire data: High preference difference	1022	265502.34	0.00001454**
Style attribute (baseline)	1023	300996.97	0.8187
Style attribute: Low preference difference	1023	343584.96	0.005922**
Style attribute: High preference difference	1023	255952.79	0.002262**
Style + price attribute	959	324271.00	0.02256*
Style + price attribute: Low preference difference	959	356355.07	0.03479*
Style + price attribute: High preference difference	959	290022.63	0.01677*
Style + category attribute	984	284767.52	0.0391*
Style + category attribute: Low preference difference	984	296520.31	0.3658
Style + category attribute: High preference difference	984	252827.24	0.04475*

4.1.3 Decision conflict

As described in the literature review, subjects experience decision conflict when choosing between options given to them. The topic has already been looked into through response time and Polygon area measurements, but to gain even better understanding, the preference difference ratings were also researched. In a similar manner to the studies by Philiastides & Ratcliff (2013) and Enax et al. (2016), the empirical probability of the respondent choosing the option they had indicated to prefer more in the first section of the study, was analyzed. As the Likert scale in the preference questions of the survey was a 7-point scale (-3 extremely unlikely, 3 extremely likely), the preference difference ratings ranged from -6 to 6, instead of just absolute numbers as with the analyses above. Here, the negative numbers indicate that option 2 was preferred over option 1, and the other way around for positive numbers. The bigger the number in each direction, the bigger the difference between the preferences were. In the middle at 0 is the "indecision point" where there was no difference between the ratings. In addition to the preference difference rating variable, a new variable, "more preferred choice", was created which included an encoding of 1 and 0 based which option was chosen, where 1 indicated option 1 and 0 indicated option 2.

After creating this variable, an analysis was conducted on how big of a proportion of choices were made in accordance with the initial preference stated. So, in other words, how often the respondent chose the shoe they had preferred more, within each of the different attributes and on average throughout the data. This was done by looking into the average value of the more preferred choice -variable for each of the difference ratings which are visualized in Figure 6. From the graph, it can be clearly seen that at each end of the line, the respondents have always chosen the option they preferred, while in the middle it is quite even between the options. It should be noted though that the bins for each difference rating varied in size, where the bins at the end points had less responses than the bins in the middle. In Figure 6, the blue line represents the proportions for the entire data. Each of the chosen attributes, style, price and category, are visualized with the dashed lines in red, yellow and green respectively. The shape of the line is similar to the ones found in previous literature (Philiastides & Ratcliff, 2013; Enax et al., 2016), although there is always some noise in the data that creates minor differences, but the overall trend is the similar. However, it should be noted that, for example, in their paper, Philiastides & Ratcliff (2013) studied their attributes in regards to a listed rating by the responded, due to which here the different attributes do not show as big differences among each other. But as said, the trend of the graph is still the same.

In addition to the graph, a linear regression analysis was conducted to the entire data, as well as each of the attributes. A significant effect was found between the more preferred choice -variable and the preference difference rating (estimate (standard error, SE): 0.123 (0.003); $p < 0.01$, intercept: 0.51). For tasks with style attribute only, the results were very similar as with the entire data (estimate (SE): 0.127 (0.004); $p < 0.01$, intercept: 0.49). For the price and category attributes, the regressions were done for both option 1 and option 2 separately due to the shape of the data and hence, both will be presented here. However, very minor differences appear between them within both attributes. For option 1 price, significant effect was found (estimate (SE): -0.09 (0.02); $p < 0.01$, intercept: 0.55) as well as for option 2 price (estimate (SE): 0.12 (0.02); $p < 0.01$, intercept: 0.45). However, neither operated as moderators to the difference rating ($p > 0.05$). For category attribute, neither option 1 or option 2 had a significant effect on the difference rating (Option 1: estimate (SE): 0.02 (0.03); $p > 0.05$, intercept: 0.51 / Option 2: estimate (SE): -0.001054 (0.026); $p > 0.05$, intercept: 0.52). In addition, neither performed as moderators to the difference rating ($p > 0.05$).

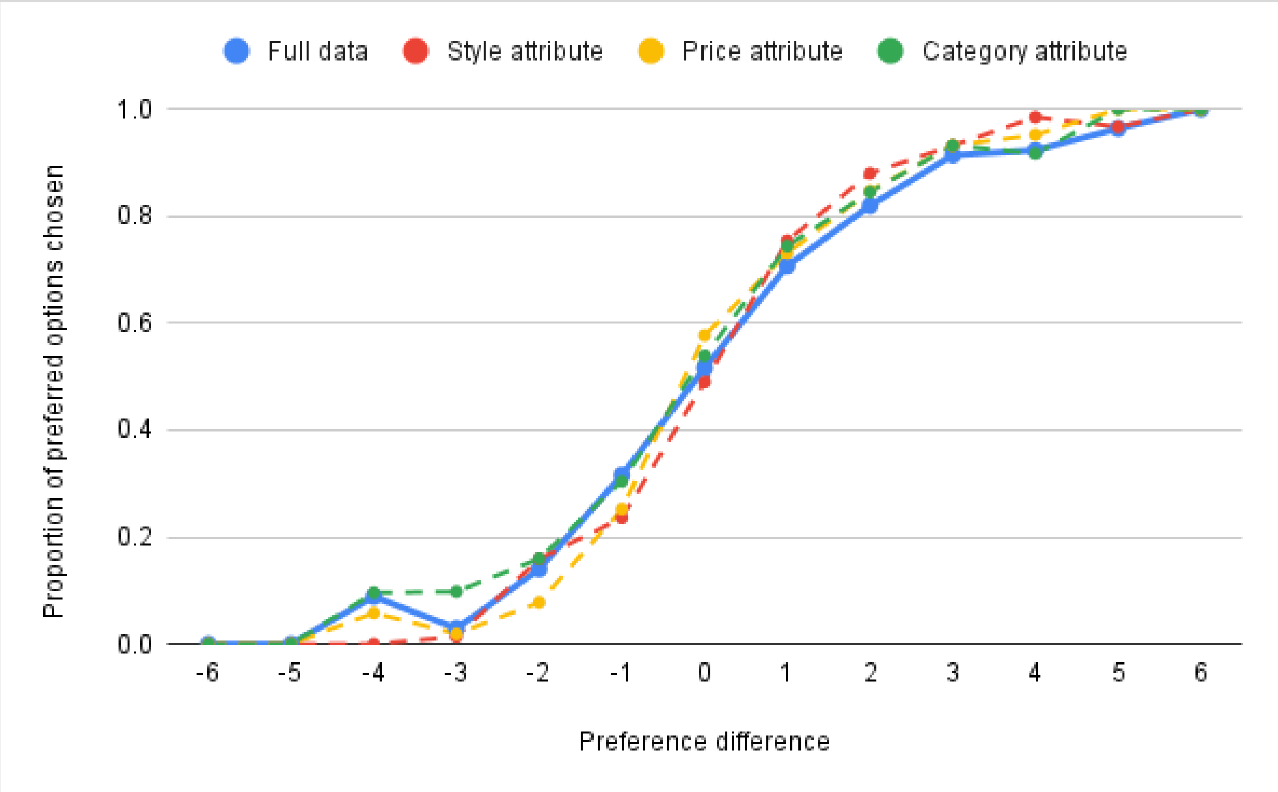

Figure 6. Empirical probability of preferred choice as a function of preference difference

In addition to comparing the preference difference rating with the more preferred choice -variable, a similar graph was also created for response time and Polygon area against the difference rating which are visualized in Figure 7 and 8. These graphs were also inspired by the research of Philiastides & Ratcliff (2013), who did a similar one with mean reaction time in comparison to item difference rating. For this research, the idea was to see how the response time altered between each the difference ratings, as well as the Polygon area size. Similar to the results of Philiastides & Ratcliff (ibid), the response time graph creates a concave line where the highest response times appear around the indecision point where the difference rating is 0 or close to it. Although there is some noise between the attributes, again the overall trend seems to follow a similar path as in previous research. However, interestingly there was some difference between the mean times for the end points -6 and

6, where for -6, the response time dropped to 3.56 seconds while for 6 only to 4.89 seconds within the entire data. For the Polygon area, the overall trend with the entire data was similar to response time. Here as well the end point -6 dropped lower than the other end at 6, where the Polygon area sizes were 146025.88 and 268693.52 respectively. However, there seems to be a bit more variation between the attribute subsets than with response time.

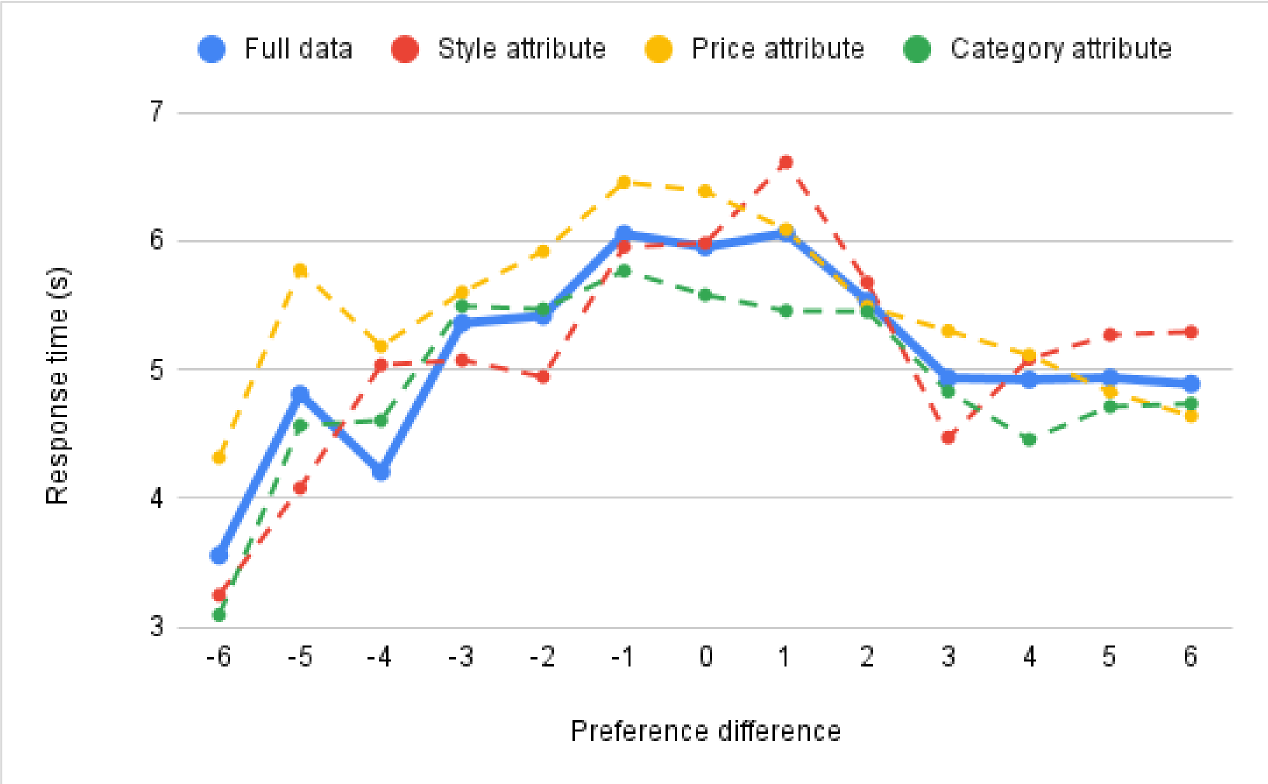

Figure 7. Response time as a function of preference difference

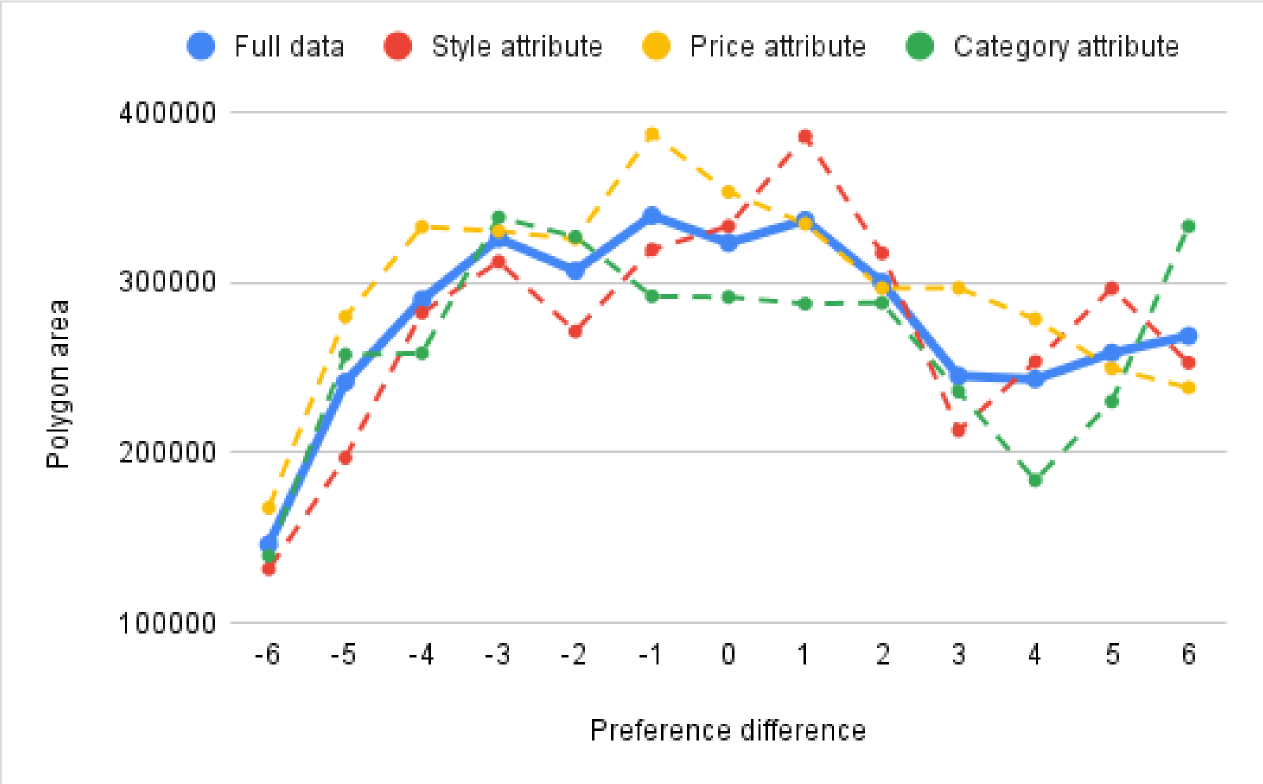

Figure 8. Polygon area as a function of preference difference

To continue the analysis on decision conflict one step further, an additional linear regression analysis was done to investigate the strength and effect of the different attributes. As the style attribute had no specific variable, such as normal price / -40% discount with price, no regression was done to it and we can already see, for example, from Figure 6 how the initial preference impacted the choice made. However, both price and category attributes were investigated to see if they had a statistically significant effect on the choice. As before, since the price and category attribute's data are in two variables each for option 1 and option 2, both will be presented here. First, the effect of price on the choice was analyzed. According to the linear regression analysis, both variables for price were found to have a significant effect (Option 1: estimate (SE): 0.13 (0.03); $p < 0.01$ / Option 2: -0.11 (0.03); $p < 0.01$ / intercept (for both): 1.47). This means that when option 1 changes from discount to normal price, options 2 is chosen 13% more than option 1 regardless if option 2 is discounted or not, while similarly, if option 2 changes from discount to normal price, option 2 is chosen 11% less than option 1. This suggest that if the option is not on sale, the respondent is more tempted to choose the other option.

In addition, the price attribute was also controlled for respondents' preference by adding that preference as a covariate to the model, in a similar manner as Enax et al.

(2016) did in their research. This linear regression model showed that even when the preference was controlled, the price attribute is still part of explaining the choice made (Price, option 1: estimate (SE): 0.08 (0.02); $p < 0.01$ / Price, option 2: estimate (SE): -0.12 (0.02); $p < 0.01$ / Preference: estimate (SE): -0.12 (0.005); $p < 0.01$ / Intercept (for all): 1.51). Finally, a logistic regression was conducted to get a further understanding of the effect of the shoe having the -40% discount -price category versus normal price. To do this, the estimates from the regression were transformed to odds ratios. For price option 1, the estimate was 0.51 ($p < 0.01$) that resulted to 1.67 odds ratio, which meant that when the price category changed from discount to normal price, the probability of the choice being option 2 than option 1, was 1.67 times more likely. For option 2, the estimate was -0.47 ($p < 0.01$) that gave an odds ratio of 0.63, which meant that when the price changed from discount to normal price, the odds of the respondent choosing option 2 was 0.63 times in comparison to choosing option 1, meaning there were more choices for option 1.

Next, the same linear regression analysis was done for the category attribute. From this, it was found that option 1 was not significant (estimate (SE): 0.007 (0.03); $p > 0.05$; intercept (for both): 1.52), while option 2 was (estimate (SE): -0.07 (0.03); $p < 0.05$). However, since they both included the same options in randomized order, it is probably due to randomness that one was found significant and the other one was not. It was also tested if controlling the category attribute with preference had any effect. According to this model, both option 1 and 2 for category were also not significant ($p > 0.05$) while preference was. Therefore, category did not have a significant effect on the choice and logistic regression was not performed.

5 Discussion

In this section, a further analysis and discussion on how the current research and its results relate and compare to the previous literature is conducted with a focus on decision conflict. Although, as mentioned, the analysis and discussion are exploratory in nature, its findings are also reflected with previous literature where applicable. As the current study should be viewed as a pilot for the research method, the use of the said method will also be evaluated, both for this research as well as an alternative to other research methods.

5.1 General discussion

In the current research, it was investigated how the level of decision conflict affects the decision making by tracing the its process with an online survey presented in section 3. The analysis in section 4 begun with an investigation into response time and Polygon area measurements in relation to research question 2 about the impact of decision conflict on the decision-making process. Both measurements were able to show that when comparing the results between low preference difference and high preference difference tasks, the response time and the polygon area increased with low difference tasks. In a similar manner as in the research by Philiastides et al. (2013), this preference difference scale acted as measure of task difficulty in relation to decision conflict: the decisions between very positive and very negative ratings were believed to be easier than between those with difference rating that was very close to each other or even equal. As said, this assumption was supported by both response time and Polygon area results for each of the attributes as well as for the entire dataset.

It should be noted though that since the research setting in this study differ from the ones discussed in existing literature, such as Philiastides et al. (2013), the actual numbers for both response time and Polygon area measurements cannot be directly compared but similar trends can be looked for. This was, for example, visible in Figure 7 that visualized the response time in comparison to the difference ratings. Philiastides et al. (ibid) found in their research a similar trend with response time (they didn't have Polygon area measurement) where the response times were shorter when the difference rating was high and longer around the indecision point (close or equal preference). In the same study, Philiastides et al. found that the participants slowed down in their decision-making process when there was incongruity between the showed item's level of preference rating and branding preference – an attribute they measured – which could be compared to situation in this research where the respondent encountered a decision task where, for example, a highly liked shoe with normal price was paired with a highly disliked with discounted price. However, this was not specifically analyzed in the current research.

Previous literature, such as the study by Stillman et al. (2020), suggest that measuring just response time isn't enough as it is more influenced by motor latencies than for example AUC measurement, which operate in a similar manner as the Polygon area measurement in this research. They found that AUC performed at least as good as response time and in some measurements even outperformed it (ibid). Koop & Johnson (2011) also discuss the fact that response time merely measures how long it takes for the subject to make a decision, but not the nature of the thought process that occurs during that time. Hence, it felt important to measure Polygon area in addition to response time, in order to have a fuller understanding. In relation to research question 2, in the current research, both response time and Polygon area showed very similar results that supported the assumption of decision conflict in the decision making. Since most of the scenarios showed statistically significant change for the low difference tasks and the high difference tasks, it could be said that having to choose between very similar options makes the decision harder and hence, requires more time and pondering between the options. Although they measured decisions under risk, Stillman et al. (2020) also found that the more similar the respondent's subjective value for the given options was, the less direct the mouse cursor trajectories were – which meant a higher AUC score – even when they controlled the results for response time.

The aim of research question 1 was to find out if the data collected from the process tracing survey could be used to measure decision conflict. Since in the survey the respondent was not able to properly see both options without scrolling between them, it could be assumed that the metric is quite sensitive to measuring decision conflict, maybe even more than AUC measurement. The analysis also looked into the weights and importance of the different attributes as well as the preferences in relation to research question 1. Consumer choice has been tested in various ways by manipulating different attributes, such as brand labels (Philiastides et al., 2013), visual saliency (Enax et al., 2016; Milosavljevic et al., 2012) or even the number of attributes available (Fasolo et al., 2007), while in this research, the chosen attributes were style, price and product category. As visualized in Figure 6, the proportion of choosing the shoe that the respondent had stated to prefer more increased when the difference ratings clearly differed, all the way to the point of always choosing the more preferred shoe at end points of the difference rating scale while in the middle, around the indecision point, the choice was quite 50/50 between the options with all of the attribute subsets as well as the entire data. A similar trend was also found in the research by Philiastides et al. (2013) with proportion of more preferred brand choices in comparison with item difference rating and Enax et al. (2016) with proportion of healthy choices in comparison with taste preferences. In the current research, it was also found that the price attribute did have a significant effect on the respondent's choice even when the subjective preference towards the shoe was controlled for, but it seems that preference still had a strong influence as shown by the graphs in Figure 6 that only presented very minor differences between the attributes. Also, as Table 1 visualizes, the price attribute affected the response time measurements the most while category attribute the least. However, the category attribute questions were the last ones in the survey and it could be possible that at that point the respondents were already more familiar with the task and were therefore able to make decisions faster. The same trend was also visible with motor responses, as the Polygon area in general was the highest with price attribute tasks and lowest with category attribute tasks. The results suggest that price had the biggest effect on the consumer decision making out of the attributes present in this research and within the price attribute, normal price seemed to drive the respondents more towards choosing the other option.

5.2 Method evaluation

As mentioned, the current research method has not been used that much in decision process tracing studies for consumer behavior before and therefore, sheds an interesting insight to its use. The current, popular research methods in process racing – especially mouse-tracking and eye-tracking – provide great research methods result-wise but possess shortcomings with flexibility and affordability. The work of Stillman et al. (2020) was able to demonstrate that decision conflict in fact exists in a continuum, not as a binary variable which either is there or is not, where the level of conflict is easily affected by the setting of the decision. Therefore, measuring conflict with process tracing instead of just choice outcomes is important. The motor responses literature also speaks for this. As Koop & Johnson (2011) state, "the research on response dynamics captures the continuous, online processing of information as it is revealed in the subject's response" (p. 751) which they believe is the way to go when talking about process tracing. Stillman et al. (2020) also found that tracing mouse trajectories (i.e. motor responses) can provide unique insights into the subject's preferences, even if the final decision would remain the same. Hence, there seems to be no question in the decision-making field that the existing methods are good methods to gain a better understanding of the decision-making process, which also gives a green light to experimenting with different methods – such as this one – that measure similar things.

All in all, this method seemed to work well for its purpose and succeeded especially in its ability to easily create a survey that works in all devices and all browsers without needing to install any software, while still being able to track similar measures as other process tracing methods do, such as response times. This possibility made it easier to collect responses to the survey, especially in the situation where there was no possibility to compensate the respondents for their efforts. Although there were some technical problems with few respondents, these are likely things that could be fixed in future research for example by having a test group first and by testing with, for example, smaller images. Having the survey done on hand-held devices also seemed logical, as online shopping and hence, consumer decisions, have greatly shifted to mobile settings in the recent years. This offered therefore a more realistic setting. However, the survey of course could have been done in a way that would more resemble an actual online store.

In relation to research question 1, the data gathered from the survey was able to give a possibility to look into decision conflict from various angles. In their research, Stillman et al. (2020) discuss how they perceive mouse tracking to possess three primary benefits over other process tracing methods which were also discussed in the literature review section of this research: "First, it provides a face-valid, readily interpretable dynamic assessment of choice. Second, it is easily accessible to researchers and practitioners alike, requiring no expensive equipment or extensive training to use. Third, the approach of measuring cursor movements is scalable beyond the lab, allowing researchers to covertly assess conflict outside of the laboratory." (p. 31739). After conducting the current research, it seems that using oTree for this type of survey design offers these same benefits. First, getting the data from oTree was easy and, for example, visualizing the scrolling movements from each response could be done from the x- and y-coordinates the survey collected. Second, oTree is a free, open-source platform that can be accessed by anyone and that offer creating interactive experiments without installing any software i.e. the platform is fully browser-based (Chen et al., 2016). Third, having the feature of being a browser-based platform, surveys and other experiments can be easily deployed beyond laboratory settings, without any major restrictions on scalability. In addition, as the experiments in oTree are created by programming the wanted features (the platform provides a lot of templates that decreases the need to write code yourself), it allows for creating basically any type experiment that can run on a web browser which could give way for even further possibilities than the ones mentioned above. Stillman et al. (2020) also points out in their research that often in real life, decisions have several options and not just two as in many studies, but the interpretation of mouse movements in such setups is possible, but not as straightforward. Even though the current research also had just two options per decision task and cannot therefore with confidence say that oTree would be definite solutions to this matter, it is definitely something that could be tested in the future. All in all, the conclusions and findings from this research suggest support for research question 1 as the method seemed to operate as valid way for measuring decision conflict and the strength of the attributes.

In addition, in the current age of online shopping with touch-screen mobile devices, it is important to consider if studying mouse cursor movements to understand decision conflict (or other decision process elements) is as relevant as it once were, and at least its' limitations should be taken into account. Stillman et al. (2020) also discuss the same thought that measuring conflict should not be limited to mouse cursor movements and could in fact explore other measurements, such as scrolling, which is exactly what was done in the current research. To continue, they state that “with the rise of mobile devices as the primary way many people access the internet, future research should investigate motor indicators of conflict that do not require a mouse” (ibid, p. 31745). This is something that should definitely gain more attention in the future, in order to the decision-making field to keep up with current consumer decision-making environments.

6 Conclusion

The final section of the thesis concludes the main points of the research and provides a discussion on its implications to real life. It also provides explanation of the limitations of the current research, accompanied by suggestions for the future studies.

The current research conducted a pilot for the use of oTree, an open-source platform (Chen et al., 2016), to program a survey that operates as process tracing method in consumer decision making tasks. The aim of the study was two-folded: first, it wanted to test how this method would work for consumer decision making situations as an alternative to other process tracing methods, and second, it wanted to see how the measurements from the survey were able to shed light to decision conflict and its effect on the consumer decision-making process. In section 1 the background and goals of the study were presented. Section 2 looked through the existing literature in the field that was found relevant to the current topic. In section 3, the used research method was presented as well as the survey itself in depth. Section 4 presented the data and findings from the survey, followed by section 5 that further explored these findings and discussed them in relation to previous findings in the field. Section 5 also entailed an evaluation of the research method.

The main findings of the research were that the method was in fact able to operate as an alternative to other process tracing methods in consumer decision making situation, especially to mouse-tracking as the current method also relies on motor responses by hands. It was also able to sensitively visualize the level of conflict during the decision-making process. The research was able to show that the decision conflict increased in all scenarios when the preference difference was low between the options compared to high preference difference tasks, in terms of both response time and Polygon area. Both measurements also showcased that from the attributes, price had a significant effect on the choice, while category attribute did not. It was also visible in the data that the proportion of choosing the shoe the respondent has reported to prefer more, clearly increased when the difference rating was either highly positive or highly negative, while around the indecision point where the difference was either very low or even equal, decisions were quite even between the options. All in all, the research method was found applicable for the research setting and could even provide more current insights to today’s consumers’ behavior by tracking scrolling instead of mouse cursor movements.

6.1 Implications

The findings of this study could provide implications for both researchers and businesses. For researchers, this provides a pilot for the use of another process tracing method that has not been widely used yet within decision-making field, especially in regards to consumer decisions. In the current shift to online shopping with mobile devices, it is important to look further from just measuring mouse cursor movements to measuring the scrolling process of the decision maker. Online shopping environment also differs from the physical store by having basically unlimited possibilities for information presentation which can even lead to information overload. Hence, it would be important to better understand how consumers are able to navigate in this environment. The current research also provides understanding of decision conflict in consumer decisions and how that can be measured with the presented method. Although the implications may lie more heavily on the researcher-side, also businesses could take the findings into consideration. Especially online stores, where the level of decision conflict could rise with the increase of available options, could benefit from a better understanding of decision conflict and for example, on the relationship between subjective preference and price. However, further, practical implications would require more research on the topic.

6.2 Limitations & further research

While offering good insights to its topic, the current research also possessed some limitations that should be acknowledged and considered for a possibility to improve them in future studies. First, as said many times, the study operated as pilot for the research method and therefore, it would of course need more testing to become more reliable and to offer points of comparison. Not only could it be used in similar certain situations with consumer decision making with, for example, different products or attributes, it could also be used in different decision-making settings, such as decisions under risk as gamble or lottery. Also, even though the sample size was acceptable for this research, especially the testing of the research method could benefit from a bigger sample size that would provide a larger dataset. The study could also be tested with different, more niche samples, for example, based on demographics. Also, it should be noted that as the survey was done online by the respondents on their own time, the response setting was not controlled which could affect their concentration on the responses.

In addition to the analysis done in this research, it could be taken a step further by analyzing the respondents' individual data, as for example, Philiastides et al. (2013) did. The respondents could be also asked for their subjective importance of the different, chosen attributes, that could be further used in the individual behavior analysis. Also, it was not in the scope of this research to have a deep dive into the respondents' behavior with different combinations of the attributes, such as comparing situations where the preference of one shoe was high, but it was also normal priced and the preference of other was low, but with a discount price category. Looking further into these relationships with the attributes could provide interesting insights.

As many studies in the decision-making field, the current research method could be accompanied by another research method, such as eye-tracking, which could provide further comparability for the method. Naturally, also other analyses could have been done with the current data or at least with a similar study, such as the drift-diffusion model (DDM) that is popular within decision making studies (for example Enax et al.2016; Philiastides et al., 2013). A suggestion for future studies in general would be to also look more into scrolling behavior of consumers, as consumers are shifting more towards shopping online on mobile devices. In order for the decision-making field to stay up-to-date with consumer behavior in online settings, this important and significant change in the decision-making environment should be considered.

7 References

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, 25(3), 187–217. https://doi.org/10.1086/209535

Boardman, R., McCormick, H., & Henninger, C. E. (2022). Exploring attention on a retailer's homepage: an eye-tracking & qualitative research study. *Behaviour & Information Technology*, 1–17. https://doi.org/10.1080/0144929X.2022.2059396

Bulling A. & Wedel, M. (2019). Pervasive eye-tracking for real-world consumer behavior analysis. In: Schulte-Mecklenbeck, M., Kuehberger, A. & Johnson, J. G. (eds.) *A Handbook of Process Tracing Methods for Decision Research: 2nd Edition* (2^{nd} Edition). Routledge, New York, NY.

Chatterjee, S., & Heath, T. B. (1996). Conflict and Loss Aversion in Multiattribute Choice: The Effects of Trade-Off Size and Reference Dependence on Decision Difficulty. *Organizational Behavior and Human Decision Processes*, 67(2), 144–155.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97. https://doi.org/10.1016/J.JBEF.2015.12.001

Cheng, J., & González-Vallejo, C. (2018). Unpacking decision difficulty: Testing action dynamics in Intertemporal, gamble, and consumer choices. *Acta Psychologica*, 190, 199–216. https://doi.org/10.1016/j.actpsy.2018.08.002

Chocarro, R., Cortiñas, M., & Villanueva, A. (2022). Attention to product images in an online retailing store: An eye-tracking study considering consumer goals and type of product. *Journal of Electronic Commerce Research*, 23(4), 257–281.

Cohen, A. L., Kang, N., & Leise, T. L. (2017). Multi-attribute, multi-alternative models of choice: Choice, reaction time, and process tracing. *Cognitive Psychology*, 98, 45–72. https://doi.org/10.1016/j.cogpsych.2017.08.001

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, 35(1), 15–28.

Dhar, R., & Wertenbroch, K. (2000). Consumer choice between hedonic and utilitarian goods. *Journal of Marketing Research*, 37(1), 60–71. https://doi.org/10.1509/jmkr.37.1.60.18718

Dyer, J.S. (2016). Multiattribute Utility Theory (MAUT). In: Greco, S., Ehrgott, M. & Figueira, J. (eds.) *Multiple Criteria Decision Analysis*. International Series in Operations Research & Management Science, vol 233. Springer, New York, NY. https://doi.org/10.1007/978-1-4939-3094-4_8

Enax, L., Krajbich, I., & Weber, B. (2016). Salient nutrition labels increase the integration of health attributes in food decision-making. *Judgment and Decision Making*, 11(5), 460–471.

Fasolo, B., McClelland, G. H., & Todd, P. M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory*, 7(1), 13–26. https://doi.org/10.1177/1470593107073842

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10, 171-178.

Gidlöf, K., Anikin, A., Lingonblad, M., & Wallin, A. (2017). Looking is buying. How visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. *Appetite*, 116, 29–38. http://dx.doi.org/10.1016/j.appet.2017.04.020

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, 18(3), 384–401. https://doi.org/10.1177/1368430214538325

Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment and Decision Making*, 6(8), 750–758.

Krucien, N., Ryan, M., & Hermens, F. (2017). Visual attention in multi-attributes choices: What can eye-tracking tell us? *Journal of Economic Behavior & Organization*, 135, 251–267. https://doi.org/10.1016/j.jebo.2017.01.018

Lee, S. H., Liang, Y., Chen, Y., Mahdi, A., & Qin, J. (2021). Can consumers' visual attention be predictable? A saliency modelling-based approach on fashion advertisements. *International Journal of Fashion Design, Technology and Education*, 14(3), 253–262. https://doi.org/10.1080/17543266.2021.1925354

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, 51, 1085–1101. https://doi.org/10.3758/s13428-018-01194-x

March, D. S., & Gaertner, L. (2021). A method for estimating the time of initiating correct categorization in mouse-tracking. *Behavior Research Methods*, 53, 2439–2449. https://doi.org/10.3758/s13428-021-01575-9

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, 22(1), 67–74. https://doi.org/10.1016/j.jcps.2011.10.002

Nielsen Norman Group (2021). How Many Participants for Quantitative Usability Studies: A Summary of Sample-Size Recommendations. Available from: https://www.nngroup.com/articles/summary-quant-sample-sizes/ [Accessed 6.7.2023]

Niza Braga, J., & Jacinto, S. (2022). Effortless online shopping? How online shopping contexts prime heuristic processing. *Journal of Consumer Behaviour*, 21(4), 743–755. https://doi.org/10.1002/cb.2032

Nunan, D., Birks, D. F., & Malhotra, N. K. (2020). *Marketing Research: Applied Insight* (6th Edition), Pearson: UK.

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica,* 144, 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Philiastides, M. G., & Ratcliff, R. (2013). Influence of Branding on Preference-Based Decision Making. *Psychological Science*, 24(7), 1208–1215. https://doi.org/10.1177/0956797612470701

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

Stillerman, B. S. & Freeman, J. B. (2019). Mouse-tracking to understand real-time dynamics of social cognition. In: Schulte-Mecklenbeck, M., Kuehberger, A. & Johnson, J. G. (eds.) *A Handbook of Process Tracing Methods for Decision Research: 2nd Edition* (2^{nd} Edition). Routledge, New York, NY.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How Mouse-tracking Can Advance Social Cognitive Theory. *Trends in Cognitive Sciences*, 22(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Stillman, P. E., Krajbich, I., & Ferguson, M. J. (2020). Using dynamic monitoring of choices to predict and understand risk preferences. *Proceedings of the National Academy of Sciences of the United States of America*, 117(50), 31738–31747. https://doi.org/10.1073/pnas.2010056117

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157, 1124–1131. https://doi.org/10.1126/SCIENCE.185.4157.1124

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.

Wang, Q., Yang, Y., Wang, Q., & Ma, Q. (2014). The effect of human image in B2C website design: An eye-tracking study. *Enterprise Information Systems*, 8(5), 582–605. https://doi.org/10.1080/17517575.2014.925585

Zuschke, N. (2020). An analysis of process-tracing research on consumer decision-making. *Journal of Business Research*, 111, 305–320. https://doi.org/10.1016/j.jbusres.2019.01.028