Eindhoven University of Technology

MASTER

Investigating the effect of using personalized critiques in a decision-making tool

Frederix, A.P.J.

Award date:
2021

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 07. Nov. 2026

Department of Industrial Engineering Innovation Sciences
Human-Technology Interaction Research Group

Investigating the effect of using personalized critiques in a decision-making tool

Master Thesis

A.P.J. (Anne) Frederix
1396676

Supervisors:
First supervisor: Dr. ir. M.C. Willemsen, Eindhoven University of Technology
Second supervisor: Dr. ir. M.A. Conijn, Eindhoven University of Technology
Company Supervisor: Dr. P. van der Aa, Den Haag The Consumentenbond

Final version
Eindhoven, 2nd of June, 2021

Abstract

Over the last few decades, e-commerce has grown rapidly in popularity. This shift to an online environment has many benefits. However a downside to the growing number of online retailing is that the choice process can become more complex. Therefore many businesses are trying to offer some decision support. Most online (shopping) environments use non-compensatory strategies to help consumers in finding a product which might lead to sub-optimal results because trade-offs are being ignored. There are well-supported findings that compensatory-based support is normatively better. The Consumentenbond is one of the few organizations that is already trying to support users using compensatory strategies to calculate the product's utility. Possessing such knowledge allows the Consumentenbond to provide reliable advice. However, their decision-making tool (called the "Keuzehulp") still uses a non-compensatory strategy. To overcome this problem, the present study proposed to develop a decision support tool that uses a compensatory approach to give normatively better support. Furthermore, researchers have argued that not everyone might appreciate the same type of support. Therefore, we investigated two personal characteristics which might influence the perception and experience of the Keuzehulp.

Critiquing has been found to guide consumers in efficiently finding their ideal product, which is particularly meaningful when consumers are searching for durable products. To investigate if adding a personalized critiquing-based element to the Keuzehulp would improve users' perception and experience with the support tool, a new Keuzehulp was developed. Subjective perceptions and experiences were measured after participants interacted with the Keuzehulp by questionnaires measuring choice satisfaction, perceived guidance, and perceived usefulness. The two personal characteristics we looked into were maximization tendency; People who maximize their choice (maximizers) vs people who are already satisfied with an option that is "good enough" (satisficers) and level of indecisiveness (indecisive vs decisive people).

A study was run with a total of 434 participants and showed that people are more satisfied with their choice when using the newly developed Keuzehulp. In addition, this study provides evidence that personal characteristics influence users' perception of the Keuzehulp. Which is in line with earlier research. The findings suggest that users' perceived guidance is moderated by their maximization tendency. Subsequently, we found that maximizers had more perceived guidance which in turn resulted in more choice satisfaction. Suggesting that it would be beneficial to create a Keuzehulp that tailors to the users' maximization tendency.

Acknowledgements

I would first like to express my deepest gratitude and thanks to my mentor and guide, Dr. ir. Martijn Willemsen for his continuous support and mentorship throughout this whole process. Even though a lot of our meetings were online, his passion for this topic, and helpful insights have made this whole experience most memorable.

Secondly, I would like to thank the team at the Consumentenbond, especially my supervisor, Dr. ir. Petra van der Aa, for giving me the opportunity to contribute to improving their decision-making tool. It was such a great experience working with so many people with different fields of expertise. My supervisors at the Consumentenbond have spent a lot of time and effort helping me conduct this research. They gave me the freedom to do my research and helped me in any way they could.

I also want to give a special thanks to Danny de Zeeuw, for helping me built the two simulations. Without his help, I would not have been able to create such interactive and realistic decision-making tools. His patients, enthusiasm, and insights have made this project what it is today.

Finally, I would like to thank my friends and family for their support throughout this process, especially my friend Janelle. Even though you live on the other side of the world, you have been such an incredible support.

Table of content

Chapter 1.: Introduction 1
Chapter 2.: Theoretical Background 1
2.1. Choice process 2
2.2. Decision-making strategies 3
2.3. Going beyond “one-size-fits-all” platforms 4
2.4. Types of critiquing based recommendations 6
2.5. The Consumentenbond 8
2.6. New Keuzehulp 10
2.7. Maximization tendency 13
2.8. Level of indecisiveness 14
Chapter 3.: Method 16
3.1. Experimental Design 16
3.2. Procedure 16
3.3. Participants 17
3.4. Measures 18
3.4.1. Maximization scale (PC) 18
3.4.2. Indecisiveness (PC) 18
3.4.3. Perceived guidance (SSA) 19
3.4.4. Choice satisfaction (EXP) 19
3.4.5. Perceived usefulness (EXP) 19
3.4.6. Behavioral measures 19
3.5. System design and implementation 19
3.5.1. User interface design 19
3.5.2. Algorithm 22
Chapter 4.: Results 25
4.1. Outliers 25
4.2. Establishing construct validity 25
4.3. Pre-experimental questionnaires 26
4.3.1. Maximization tendency 26
4.3.2. Level of indecisiveness 26
4.4. Post-experimental questionnaires 27
4.4.1. Choice satisfaction 27
4.4.2. Perceived usefulness 27
4.4.3. Perceived guidance 27
4.5. Structural model 28

4.5.1. Perceived guidance 30
4.5.2. Choice satisfaction 31
4.5.3. Moderated mediation 32
Chapter 5.: Discussion 33
5.1. Main Findings 33
5.1.1. Maximization tendency 35
5.1.2. Level of indecisiveness 36
5.2. limitations 37
5.3. Implications 38
Chapter 6.: Conclusion and Future research 39
6.1. Future research 39
6.2. Recommendations for the Consumentenbond 40
References 41
Appendix 1: Questionnaires experiment 49
Appendix 2: Decision tree Keuzehulp 50
Appendix 3: Design Keuzehulp 52
Original Keuzehulp 52
New Keuzehulp 53
Appendix 4: Preference-based algorithm 55
Appendix 5: Statistical output 59
Maximization tendency and level of indecisiveness 59
Goodness of fit SEM model 60
Covariances 61
Modification indices 61
SEM model output 61
Moderated mediation 61
Maximization 61
Level of indecisiveness 62

Chapter 1.: Introduction

Over the last two decades, e-commerce has snowballed in popularity (Dolfen et al., 2019). Many traditional Brick-and-Mortar retailers have shifted to online shopping environments, and new online platforms have emerged (e.g., the Consumentbond and Bol.com). The convenience, easy access to information, and wide variety of services/products are primary advantages of these e-commerce websites (Zhou et al., 2007). However, a downside to the growing number of online retailing is that the choice process can become more complex since consumers need to deal with a lot of information at the same time. This is forcing businesses to provide better support for consumers' decision-making. *But, how can we support consumers in making decisions?* This study aims to investigate how we can optimally support consumers in their search of finding a product that is most suitable. More specifically, this thesis will focus on improving a decision support tool that was developed by the Consumentenbond. The Consumentenbond is a non-profit organization whose primary goal is to help consumers in finding a product that suits their preferences by providing detailed information about these products. Therefore it is important for the Consumentenbond to offer good decision support to their consumers.

This thesis will start by discussing relevant literature on human decision-making and will describe the theories and findings needed to offer better decision support (**Chapter 2.**: Theoretical Background). Secondly, we will discuss the research questions and accompanying hypotheses based on earlier research and theories. Thirdly, the methodologies of the experiment used to test the decision support tool will be addressed (**Chapter 3.**: Method). Lastly, our findings will be outlined (**Chapter 4.**: Results), followed by the discussion (**Chapter 5.**: Discussion), limitations, and conclusions (**Chapter 6.**: Conclusion and Future research).

Chapter 2.: Theoretical Background

Chen and Cheng (2009) have argued that in order to have a successful online shopping environment, it is important to understand consumers' intentions and motives. The first step in understanding consumers' intentions is to know how consumers make choices. Therefore, this thesis will start by discussing relevant literature on human decision-making.

2.1. Choice process

A choice process often involves a sequential search, where a consumer inspects some of the (often large) number of candidate products available in the market to identify a product that best matches their preferences. When facing many options, the choice process often consists of two stages (Xiao & Benbasat, 2007; Harley, 2020). First, consumers start by quickly reducing a large set of alternatives to a more reasonable set and then thoroughly compare the options that are left to make an optimal decision.

According to the rational choice theory (Bettman, Luce, and Payne, 1998), decision-makers have well-defined preferences that do not depend on specific methods used to elicit those preferences (Doyle, 1999). This theory assumes that consumers can find an option that will maximize their perceived value and select an option accordingly. Bettman and James (1979) argued that this theory is incomplete in understanding the consumers' decision-making process. Decision-makers often have limitations in their capacity for processing information (Simon, 1955). These limitations include limited computational capabilities and limited working memory. For example, they might have difficulty understanding the attributes and their values.

There is extensive evidence (e.g., Bettman 1971) that consumers use heuristics when they are trying to make a choice. Heuristics are simplified rules that consumers use to handle the available information (Babutsidze, 2012). Additionally, Bettman, Luce, and Payne (1998) have argued that since consumers lack of ability to generate and maintain stable references, people often construct their preferences on the spot. Moreover, preferences are assumed to be constructed based on the task environment and contextual factors (Payne, 1982; Payne, Bettman, & Johnson, 1993; Simon,1990).

Since consumers often construct preferences on the spot, making choices can be challenging. Especially due to the rapidly growing number of products shown on online platforms and the many features that characterize these products. One way to help consumers construct their preferences and subsequently find a product is to offer decision support (Häubl and Trift, 2000). Therefore, many online retailers are incorporating decision-making tools on their websites

to assist consumers with their choice process (Grenci & Todd, 2002). There are many types of decision support tools that can be used. Xiao and Benbasat (2007) categorized the support systems into three types. The first type includes collaborative filtering and content-filtering. The second type includes needs-based and attribute-based recommendation agents. Finally, the third type includes compensatory and non-compensatory strategies. The present study will focus on compensatory and non-compensatory-based systems. In the following section, these two decision strategies will be elaborated on further.

2.2. Decision-making strategies

Most online shopping environments use non-compensatory strategies to help consumers in finding a product. In general, using non-compensatory decision support systems result in presenting options that meet predetermined cut-off points set by the decision-maker (Xiao and Benbasat, 2007). This strategy allows users to restrict the set of possible choices by providing them with lower and upper limits on various attributes. A commonly used non-compensatory strategy is a filtering-based method. A well-known filtering method is 'Elimination By Aspects' (EBA), which requires consumers to choose among a set of products during a sequential process (Bettman, Luce, and Payne, 1998). During each stage, consumers have to eliminate products that do not possess a particular attribute until a product is chosen. Using such a non-compensatory strategy allows consumers to have some control over manipulating and viewing features of interest. This method is helpful when a consumer wants to narrow down their search quickly or is unwilling to compromise on any of these attributes, and they have a clear idea about the optimal attribute settings.

Though decision-making support tools aim to help consumers make the most optimal and satisfying choice, one can question whether this is possible when non-compensatory strategies are being used. Using this type of strategy avoids confronting the conflicts inherent when making decisions and does not allow decision-makers to make trade-offs (Hogarth, 1987). For example, when someone is looking for a washing machine, one might filter out all the noisy washing machines. Only options that meet this criterion will be shown to the consumer. However, there could be a washing machine that is somewhat noisier but significantly less expensive. This might be a better option for the consumer, but they will never see this option since this product is filtered out, leaving the consumer with sub-optimal results. Therefore, employing non-compensatory strategies may not always yield options suggested by normative decision-making models (Gudigantala, Song, and Jones, 2008) since non-compensatory strategies do not consider a consumers' preferences for multiple attributes. In addition, at times, the use of this strategy may

result in the elimination of some alternatives based on the criterion set of one attribute, even though such options may be desirable on other attributes.

Research has shown that helping users select and compare options that share trade-off properties can help increase the level of decision accuracy (Pu, Chen, and Kumar, 2008). Additionally, Frisch and Clemen (1994) argued that making trade-offs is an important aspect of high-quality rational decision-making. Compensatory strategies enable users to resolve conflicting values explicitly by considering the trade-offs. When a consumer needs to choose between a large number of products, they must consider multiple attributes of available alternatives at the same time. Researchers have argued that more important attributes will significantly impact determining preferences (Jansen, Coolen, and Goetgeluk, 2011). Compensatory strategies allow a product's desirable value to compensate for an undesirable value of another. One of the best-known implementations of compensatory choice is Multi-Attribute-Utility Theory (MAUT) (Payne, Bettman, and Johnson, 1993). In its simplest form, MAUT considers the distinct values for a set of influential attributes and lets users assign weights of relative importance to these attributes. A products' utility is calculated by multiplying the separate attribute values with the users' assigned importance weights and summing them. Subsequently, the product with the highest score is selected.

Researchers have argued that compensatory strategies are more accurate than non-compensatory strategies when potential trade-offs exist between variables. In addition, Gudigantala, Song, and Jones (2008) argued that compensatory strategies show the consumers multiple attributes, which makes it easier for consumers to identify and assess the trade-offs that have to be made. Although there are well-supported findings that compensatory-based support is normatively better, only a fraction of online shopping environments uses this strategy. Furthermore, Gudigantala, Song, and Jones (2008) found that there is an absence of support using compensatory strategies that are based on consumers' own preferences. In other words, there is a lack of decision support that is based on a consumers' *personal* preferences, the support is one-size-fits-all.

2.3. Going beyond "one-size-fits-all" platforms

Most online shopping environments offer support that uses a one-size-fits-all approach and do not take into account the users' preferences. The same type of support is given to all consumers. A problem with such a one-size-fits-all approach is that they fail to consider interpersonal differences between users. Different people hold different beliefs, attitudes, and values and are motivated by different things (He, Greenberg, and Huang, 2010). Bettman, Luce, and Payne (1998) argued that people differ in the way they make decisions, subsequently, people might differ in the

way they want to be supported. Similar results were found by Cox and Dale (2001), who argued that customers differ in what products they want and how they want to be treated by a sales representative.

Researchers have proposed that a consumers' decision-making process depends on a consumers' personal characteristics (e.g., domain knowledge; Knijnenburg and Willemsen, 2009, 2010; Knijnenburg et al., 2012; Kamis and Davern, 2004). Additionally, Aksoy et al. (2006) found that users prefer recommendation systems that use a decision strategy like the one they would use themselves. Since people can have different personal characteristics, the appreciation for the type of decision-making tool might also differ. For example, one consumer might prefer to have more guidance, and the other might prefer to do it more independently. Using a personalized decision-making support tool that tailors to the different users could potentially recommend more personally relevant products. Jiang, Shang, and Liu (2010) have argued that the personalization of product information has become one of the most important factors that impact a consumer's product satisfaction and selection.

Looking at all the theories and findings that have been described above, creating a tool that uses compensatory strategies and is also personalized might facilitate our goal of improving users' decision support. Especially given that compensatory decision support provides better decision quality, satisfaction, and perceived usefulness (Gudigantala, Song, and Jones, 2008). However, Gudigantala, Song, and Jones (2008) have argued that using compensatory strategies can be challenging when choosing between many products. When more options and attributes are available, and someone is uncertain about the importance of the attributes, a consumer might become overwhelmed. Making trade-offs requires possibly accepting a loss on an attribute with potentially negative consequences. Therefore, it is important to develop a decision support tool that guides consumers in making trade-offs.

Most recommender systems use collaborative filtering algorithms that try to find connections between users. However, online shopping environments do not always have the necessary information about their users in order to use these popular techniques. In addition, online platforms like the Consumentenbond display durable products that are often not bought frequently, leading to the problem of "cold start". Conversational recommender systems are commonly used to help when no prior information about users is available. Consumers are guided through a sequence of recommendation cycles in which products are recommended based on some evolving model of the user's preferences. During each recommendation cycle, a consumer has the opportunity to provide feedback as a way to express their preferences and steer the recommendation in the direction of their desired products. One of the problems that occur is that

consumers rarely provide accurate or complete information to begin with, and this information is often contradicting and inconsistent (Bettman, Luce, and Payne, 1998).

In order to assist consumers in making accurate and confident decisions, critiquing-based recommender systems have emerged. Critiquing is an iterative approach, where a user evaluates the outputs generated by a recommender system. This allows a system to continuously update settings and provide users with updated recommendations that better represent desired outcomes. During a typical critiquing cycle, an initial set of products is presented to the user based on a user's initial preferences. Next, a user can give feedback by critiquing the set of products. Example critiques could be “larger screen” or “higher resolution”. Users can choose to critique products until they find a product that satisfies their needs (**Fig. 1**).

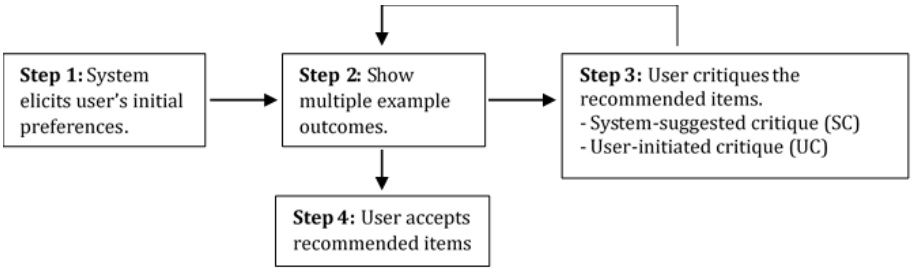

Fig. 1 A typical interaction flow of critiquing-based recommender systems (Jin et al., 2019).

Critiquing-based recommender systems have been recognized as effective recommender techniques (Chen and Pu, 2011b). Chen and Pu (2009) have argued that critiquing-based recommender systems may guide consumers in efficiently finding their ideal product, which is particularly meaningful when consumers are searching for durable products. Since many online shopping environments cannot create user profiles, critiquing could be used to discover a user’s preferences on the spot. In the past, various knowledge-based approaches have been proposed and studied, but these approaches have never reached the attention of the recommender system community (Colucci et al., 2006; Felfering, Friedrich, Jannach, and Zanker, 2006). In the upcoming section, we will go in more depth into the existing critiquing systems.

2.4. Types of critiquing based recommendations

There are two major critiquing techniques to elicit preferences, system suggested critiquing (SC) and user-initiated critiquing (UC). SCs generate a set of critiquing candidates for users to choose from. UCs focuses on stimulating users to make self-motivated critiques, where users are allowed to freely combine critiques (Chen and Pu, 2009). An example of this type of system is the Example Critiquing (EC) agent (Pu and Faltings 2004). In essence, this agent allows users to choose any

combination of feature(s) to critique. Previous work has proven that this enables users to have a higher level of control. However, users may experience difficulty getting started with UC without prior knowledge (Chen and Pu, 2012). The decision-making tool is meant to facilitate all types of consumers, and it is therefore important that the decision support tool is understandable for all their users. SC will be used in this study.

Early SC systems used unit critiquing, meaning that users express their preferences over a single feature (FindMe (Burke, 2000) and Automated Travel Agent (Linden et al., 1997)). This limits the ability of a recommendation since a user only focuses on individual features as if they were independent instead of showing them the different trade-offs. For example, when someone focuses on the price of a camera, the buyer might critique the price feature until an acceptable price has been achieved, to then only find that cameras in this region do not satisfy other requirements, for example, high resolution. The user has no choice but to start over again, wasting considerable time and effort.

An alternative way is to use compound critiques (Smyth, McGinty, Reilly, and McCarthy, 2004, September). This is a strategy where a user can critique multiple features at the same time. Compound critiques could potentially improve recommendation efficiency because they allow the recommender system to focus on multiple features within a single cycle. In addition, researchers have argued that compound critiques carry considerable explanatory power because they show users the trade-offs that have to be made (McCarthy, Reilly, McGinty, and Smyth, 2004). In the past, when compound critiques were used, the users would be presented with a fixed set of compound critiques in each recommendation cycle. A method called dynamic critiquing and its successor, incremental critiquing, both developed by Reilly, McCarthy, McGinty, and Smyth (2004, August, and December), have been proposed mainly to dynamically and automatically generate compound critiques. For example, they have created a tool that helps consumers find a camera by showing them compound critiques (e.g., “More memory & More batteries & a different strap”). Each time a consumer chooses a critique, the algorithm will save the consumers’ choices and keep track of their critiquing history. However, the main limitation of these methods is that the critique selection is purely data-driven and does not take into account a user’s preferences. It only reveals what a system can provide, but it does not consider whether a user would like the suggested critique or not.

Therefore, Chen and Pu (2007c) have developed an interface that used a more advanced critiquing technique, called Preference-based organization. It was designed to generate dynamic critiques adaptive to users' current preferences and needs. This method computes critiques with MAUT-based user preference models, which can resolve conflicting values explicitly by considering trade-offs (Chen and Pu, 2010). As discussed earlier, MAUT takes into account the separate values for a set of influential attributes and lets users assign weights of relative

importance to these attributes. The Preference-based method creates personalized critiques by iteratively improving users' importance weights of the attributes. Burke (2002) argued that providing personalized recommendations based on user's preferences improves the accuracy of a system to predict the best recommendation for the user. Subsequently, resulting in better user experience in terms of choice satisfaction (Knijnenburg, Willemsen, Gantner, Soncu, H., and Newell, 2012). In addition, Chen and Pu (2010) found that using the Preference-based critique generation algorithm achieved significantly higher critique prediction accuracy compared to using the dynamic critiquing method. Since this algorithm generates critiques based on users' preferences, this technique could potentially facilitate our goal of developing a decision support tool that is both personalized and guides consumers in making trade-offs.

To summarize, most online shopping environments are still using non-compensatory decision strategies to support consumers in making a choice. Even though there are well-supported findings that compensatory-based support is normatively better. One example is the Consumentenbond. The Consumentenbond is famous for the quality of their product advice. Due to their unique position in the market, they have the ability to optimally support their consumers. However, like most other online (shopping) environments, they are still using non-compensatory strategies to support consumers in finding a product. In the following paragraph, we will go into more depth regarding the types of support the Consumentenbond is giving and how this could potentially be improved using the theories and findings that have been discussed above.

2.5. The Consumentenbond

The Consumentenbond is a non-profit organization with over 400.000 members. For years they have tried to help consumers make good normative decisions. Since they do not sell products but focus on testing and reviewing the available products on the market, they have a unique position. They have the opportunity to give independent and detailed advice about the available products on the market. They conduct extensive product tests to gain knowledge about durable products. Possessing such knowledge and information, like product specifications, price information, and consumer reviews, allows the Consumentenbond to provide reliable advice to consumers in choosing a product.

A way they try to assist customers in finding the right product by showing them how well the products score on a set of products' attributes, which is based on extensive product tests. Subsequently, the overall product scores are calculated for each product. The Consumentenbond uses MAUT to calculate these overall test scores.

As discussed before, generally, importance weights are assigned based on a user's preferences. However, in the setup of the Consumentenbond, these weights are not based on a decision maker's personal importance but are weights that the Consumentenbond chooses and are thus the same for all customers. In **Fig. 2**, an example of a product page can be found[1], showing the products' attribute scores and their importance weights for a washing machine. A bar at the top left of **Fig. 2** shows how well the product scores compared to the best and worst product in that product domain. More additional information about the attributes can also be found; what the attributes mean, what settings were used when testing the washing machine, and their pros and cons.

Fig. 2 Example of a product page

Due to the rapidly growing number of products displayed on their website, the choice process becomes more difficult. To assist people in their search, the Consumentenbond has developed a feature called the "Product picker" to help people narrow down their search. The "Product picker" allows consumers to adjust a set of filter settings to find a product that suits their preferences, using the non-compensatory strategy, EBA.

More than most online (shopping) environments, the Consumentenbond tunes to people with a lot of domain knowledge. These "experts" can translate their needs to appropriate filter settings because they have a lot of knowledge about the products' attributes. For experts, it will

[1] Only part of the product page is shown in the figure

be relatively easy to use a feature like the "Product picker". On the other hand, Novices often struggle with the number of options and the technical descriptions of the products and therefore might need more guidance. Recently, the Consumentenbond has developed a decision tree called the "Keuzehulp". Consumers are asked a couple of questions regarding a set of important product attributes and specifications. In **Fig. 3,** the workflow of this interaction is visualized.

With just a few questions, the Consumentenbond tries to guide their member to an optimal filter adjustment to find the best option. Next, a consideration set is generated by using the non-compensatory strategy EBA.

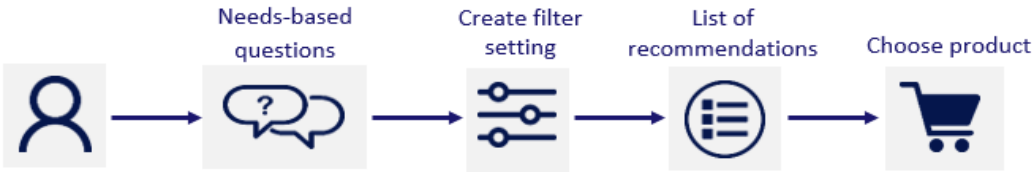

Fig. 3 Workflow Keuzehulp

The Consumentenbond is one of few organizations that is trying to give consumers advice using normative compensatory strategies. They already use compensatory strategies to calculate the products' utility. However, their Keuzehulp still employs a non-compensatory strategy to guide consumers. As discussed earlier, using non-compensatory strategies often lead to sub-optimal results since trade-offs are being ignored. In addition, the Keuzehulp uses a one-size-fits-all approach, all the questions that are being asked are the same for everyone. Since the Consumentenbond aims to offer personally relevant information and help consumers find their optimal choice, a new Keuzehulp will be developed and tested to see if it will facilitate their goals better.

2.6. New Keuzehulp

The new Keuzehulp will combine a non-compensatory and compensatory strategy. Combining the two decision-making strategies allows people to efficiently reduce a large set of alternatives to a more reasonable set and then thoroughly compare the products that are left to make an optimal decision (Xiao & Benbasat, 2007). For the non-compensatory part of the Keuzehulp, EBA will be used to create an initial consideration set (Bettman, Luce, and Payne, 1998). The reason for starting with EBA is because this technique is used in the current Keuzehulp to create the consideration set. Our goal is to expand the current Keuzehulp, not get rid of it entirely. For the compensatory part of the Keuzehulp, critiquing will be used to help consumers in making the necessary trade-offs. Since the Preference-based critiquing algorithm generates critiques based on users' preferences (Chen and Pu, 2007c), this technique will be used to personalize the Keuzehulp and potentially guide consumers in making trade-offs. As stated earlier, the

Preference-based algorithm creates personalized critiques by iteratively improving users' importance weights of the attributes. Since the Consumentenbond already uses importance weights to calculate the products' utility, these weights can be used as a starting point. Subsequently, through critiquing, these weights could iteratively become more personalized.

As discussed earlier, researchers have found that making trade-offs can often be challenging (Gudigantala, Song, and Jones, 2008). Therefore it is important to investigate if our new Keuzehulp helps facilitate our goal of guiding consumers in making trade-offs. We expect that this will indeed be the case since researchers have argued that critiques carry considerable explanatory power considering they show users the trade-offs that have to be made (McCarthy, Reilly, McGinty, and Smyth, 2004). Furthermore, Chen and Pu (2009) argued that critiquing recommender systems may guide consumers in efficiently finding their ideal product.

Moreover, the Keuzehulp aims to recommend products that satisfy its consumers. As the choice process becomes more personalized, we expect people to be more satisfied with their choice when they use the new Keuzehulp. Earlier research has found that providing personalized recommendations based on user's preferences improves the accuracy of a system to predict the best recommendations for the user (Burke, 2002). This subsequently results in a better user experience in terms of choice satisfaction (Knijnenburg, Willemsen, Gantner, Soncu, H., and Newell, 2012).

Finally, the perceived usefulness of a recommender system has been shown to positively influence a users' choice satisfaction (Alan, Kabadayi, Bakis, and Ildokuz, 2017). Additionally, Amin (2007) found that a system's perceived usefulness is an important driver of consumers' online shopping intention. Furthermore, Xia and Benbasat (2007) have argued that a personalized system increases a users' perceived usefulness of the system. Therefore, we will also investigate if users' perceived usefulness of the Keuzehulp will improve when people use the new Keuzehulp.

This leads to the following research questions and accompanying hypotheses:

RQ1: Will adding a personalized critiquing phase help improve the decision support tool by guiding users' in making trade-offs?

H1: Adding a personalized critiquing phase to the existing Keuzehulp will have a positive direct effect on users' perceived guidance in making trade-offs.

H2: Adding a personalized critiquing phase to the existing Keuzehulp will have a positive direct effect on users' choice satisfaction and perceived usefulness

H3: Perceived guidance has a positive direct effect on users' choice satisfaction and perceived usefulness.

H4: Perceived guidance mediates the effect of the type of Keuzehulp on choice satisfaction.
H5: Perceived usefulness correlates with users' choice satisfaction.

The current thesis employs the user-centric evaluation framework developed by Knijnenburg, Willemsen, Gantner, Soncu, and Newell (2012), which addresses both objective and subjective evaluations. Our goal is to improve the decision support tool and guide consumers in making trade-offs. It is, however, very difficult to measure the quality of their choices. Nevertheless, we can measure consumers' experiences and perceptions when using the new Keuzehulp.

The relevant components of the first research question and their subsequent hypotheses have been embedded in the framework. This results in the model that can be found in **Fig. 4.**

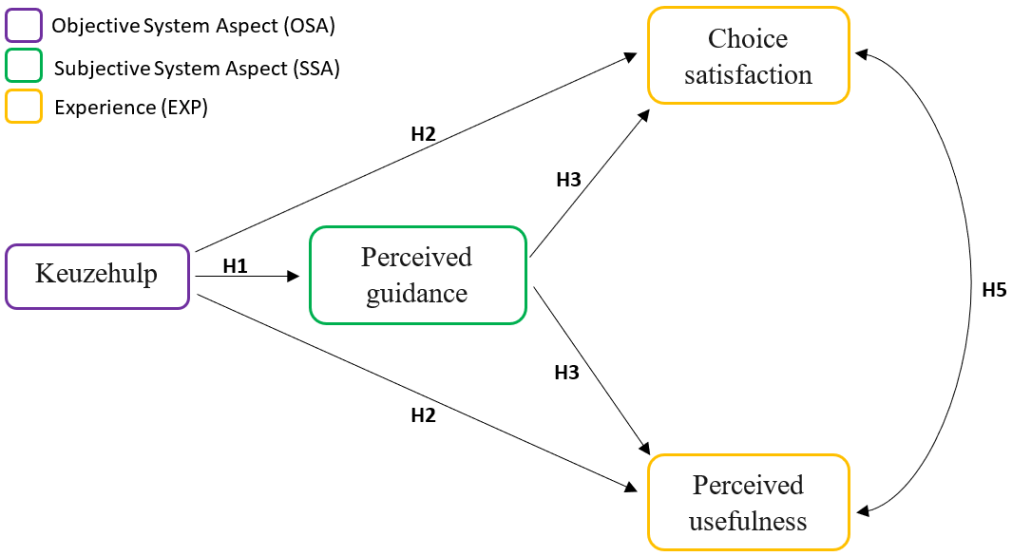

Fig. 4 The SEM fitted for RQ1.

Most existing studies regarding critiquing techniques focused on comparing different techniques and how these different techniques affect users (McCarthy, Reilly, McGinty, and Smyth, 2004; Chen, and Pu. 2007; Chen, and Pu., 2012). However, recommenders rarely consider the potential differences between users (Jin, Cai, Chen, Htun, and Verbert, 2019). Since researchers have found that people differ in how they make decisions (Bettman, Luce, and Payne, 1998), they will arguably also differ in the way they want to be guided. Therefore, it is important to investigate what factors might moderate the effect of the Keuzehulp on perceived guidance, choice satisfaction, and perceived usefulness. For example, one consumer might prefer to have more guidance, while others might prefer to search for a product more independently. Personal characteristics are factors that have been shown to influence the type of decision strategy a consumer prefers (e.g., domain knowledge; Knijnenburg and Willemsen, 2009, 2010; Knijnenburg

et al., 2012; Kamis and Davern, 2004). Additionally, Aksoy et al. (2006) found that users prefer recommendation systems that use a decision strategy like the one they would use themselves. Since people can have different personal characteristics, the appreciation for the type of decision-making tool might also differ.

Our new Keuzehulp aims to help consumers in making the most optimal choice. However, not everyone might have the same goal when looking for a product. Someone who always wants to find the most optimal product might be very satisfied with the new Keuzehulp. However, someone who is already satisfied with a product that is "good enough" might not want to put in the additional effort and could already be happy using the original Keuzehulp.

2.7. Maximization tendency

Researchers have found that people adopt different decision-making styles (Karimi, Papamichail, and Holland, 2015). A consumers'' decision-making style can be defined as the tendency to satisfice or maximize a decision. This is an individual characteristic that has been shown to be a predictor of decision-making behavior (Karimi, Papamichail, and Holland, 2015). On one extreme, you have maximizers who tend to explore most or all available alternatives before making any decisions (Jugovac, Nunes, and Jannach. 2017). Satisficers, on the other hand, evaluate options until they find one that is "good enough".

Since research has found that this characteristic might affect a consumers' decision-making behavior, it would be valuable to look into this personal characteristic. Since members of the Consumentenbond put in the extra effort to look for a product on their website instead of directly going to the websites that sell these products, we expect that a lot of these users might be maximizers. Maximizers usually engage in extensive product comparison and aspire to find the best possible option (Schwartz, 2004). Therefore maximizers might be more satisfied using the new Keuzehulp since it helps consumers make better normative decisions. On the other hand, Satisficers want to attain satisfactory results quickly (Cheek and Goebel, 2020). These individuals might already be satisfied using the original Keuzehulp and could even be unsatisfied with the new Keuzehulp since it asks for more involvement and time.

To the best of our knowledge, no earlier research has looked into the effect of consumers' maximization tendency on a decision support tool. Therefore, it would be interesting to investigate if a users' maximization tendency influences their experience and appreciation of the new Keuzehulp. And if different tools should be offered to different people, based on their level of maximization.

Schwartz et al. (2002) were the first researchers to develop a Maximization scale. One of the constructs that they used to conceptualize maximization was indecisiveness. However,

Schwartz (2016) found that indecisiveness should be conceptualized as a consequence or cause of maximization. Cheek and Goebel (2020) found that indecisiveness and maximization are two separate constructs and should not be seen as a single measure for maximization. Therefore in this study, we will look into users' maximization tendency and level of indecisiveness separately.

2.8. Level of indecisiveness

Since research has shown that indecisiveness is a consequence of maximization, it would be interesting to look into this personal characteristic. Especially since we suspect that many consumers who use the website of the Consumentenbond are maximizers.

Indecisiveness can be defined as an individual difference that describes the general tendency of experiencing difficulty when making decisions. Research has shown that people who do not have well-defined preferences often struggle with making decisions when confronted with a large assortment of choices (Chernev, 2003). Therefore it is important to investigate if our new Keuzehulp guides all types of consumers and if it depends on their level of indecisiveness. Research has shown that indecisive people tend to show a desire to seek out more information and often use this information to confirm that they are making the right choice (Ferrari and Dovidio 2000; Rassin 2007). Since the original Keuzehulp consists of only a couple of basic questions, we expect that an indecisive person might not appreciate the original Keuzehulp. Subsequently, we expect that indecisive people will appreciate the new Keuzehulp more since it shows users the trade-offs, which might help them justify their choice. However, no earlier critiquing recommender system research has looked into the influence of peoples' level of indecisiveness on users' choice satisfaction, perceived usefulness, and perceived guidance.

This paper will investigate how users' maximization tendency (maximizers vs. satisficers) and level of indecisiveness (indecisive vs. decisive people) influences their perception and interaction with the Keuzehulp. This leads to the following research questions and accompanying hypotheses:

RQ2: Do the user's Maximization tendency and their level of indecisiveness influence their perception and experience when comparing the original and new Keuzehulp?

H6: The level of maximization has a positive effect on perceived guidance when a critiquing phase is added to the Keuzehulp.

H7: The level of maximization has a positive effect on users' choice satisfaction and their perceived usefulness score when implementing a critiquing phase into the Keuzehulp.

H8: The level of indecisiveness has a positive effect on perceived guidance when a critiquing phase is added to the Keuzehulp.

H9: Implementing a critiquing phase has a positive effect on users' choice satisfaction and their perceived usefulness score for users who tend to be more indecisive.

The relevant components of both research questions and their subsequent hypotheses have been embedded in the framework. This results in the model that can be found in **Fig. 5**.

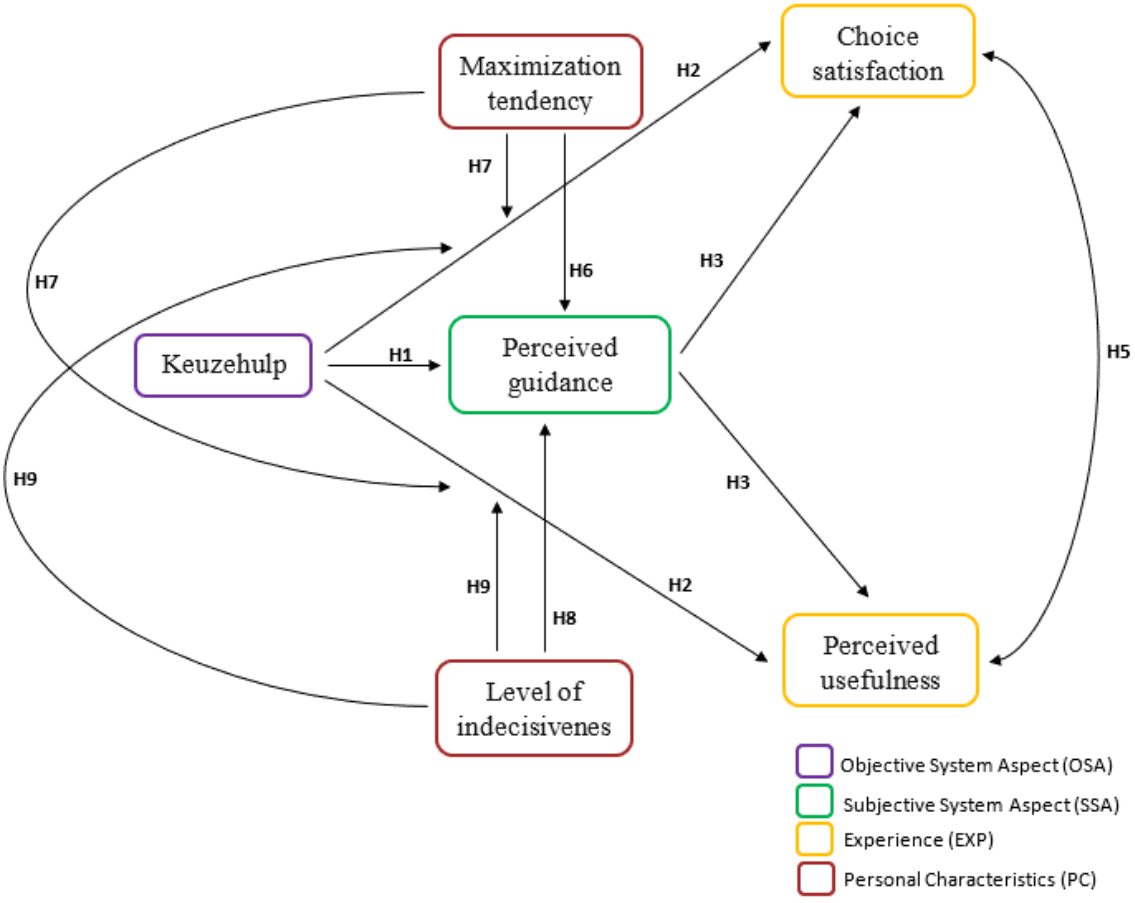

Fig. 5 The SEM fitted for this study.

Chapter 3.: Method

As outlined above, this study aimed to investigate if developing a Keuzehulp that uses both a non-compensatory (EBA) and compensatory strategy (personalized critiques) could improve the current Keuzehulp and help guide consumers in making trade-offs between attributes. Furthermore, this study investigated if users' maximization tendency and -level of indecisiveness could influence users' perception and interaction with the Keuzehulp.

3.1. Experimental Design

This study followed a between-subjects design to test the aforementioned hypotheses and to answer the research questions. The experiment took place online, where participants used one out of the two decision-making tools. The participants were randomly assigned to either the control group or the experimental group. Participants in the control group interacted with the current Keuzehulp that the Consumentenbond developed. Participants in the experimental group used the newly developed Keuzehulp. Simulations were created for the two decision-making tools and were applied for the product domain washing machines.

3.2. Procedure

The experiment consisted out of three parts; 1) answering a pre-experimental questionnaire, 2) interacting with either the original or new Keuzehulp, and 3) answering a post-experimental questionnaire (**Fig. 6**).

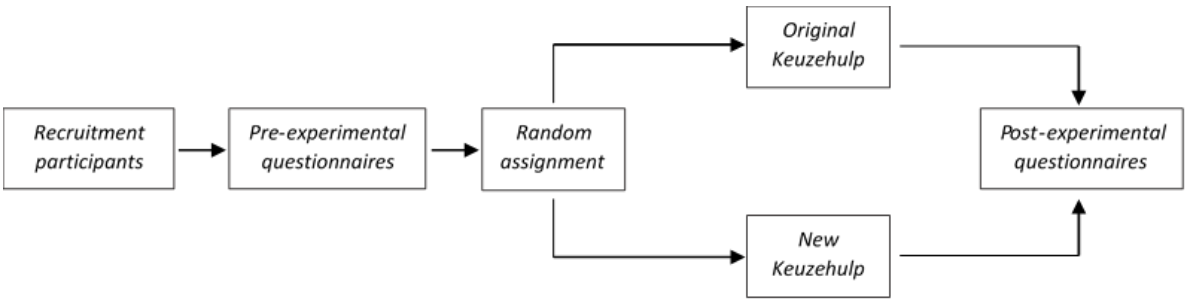

Fig. 6 Process flow diagram experiment

Participants were recruited through the panel of the Consumentenbond. They received an invitation via mail, using a "neutral" explanation to ensure that the study's true purpose was not revealed. Information was given about the different steps of the experiment and the time investment. After following the link in the email, participants were directed to a platform called "Jambo". The Consumentenbond uses this platform to send out questionnaires; hence, the same

platform was used in this study. Participants were given several pre-experimental questionnaires measuring their maximization tendency and level of indecisiveness. Next, they were given a step-by-step explanation of the simulation of the Keuzehulp. Participants were then directed via a unique link to the Keuzehulp to keep track of the users' clicking behavior. The simulation ended when a product was chosen. Finally, a post-experimental questionnaire was presented to the participants, measuring choice satisfaction, perceived usefulness, and perceived guidance. The experiment took, on average, fifteen minutes and was conducted in Dutch. The study ran for 20 days and took place from the 9th of March until the 29th of March.

3.3. Participants

In total, 434 participants were recruited in the study (194 females and 239 males, one unknown), of which the ages ranged from 25 to 45 years old (M = 37.7, STD= 5.0). This age group was chosen as requested by the Consumentenbond. The participants were all part of the panel of the Consumentenbond and would occasionally take part in their studies. Participants were not necessarily looking for a new washing machine. Recommender system research manipulations often have small effects, causing differences of around 0.2-0.3 standard deviations in the dependent variable (Knijnenburg and Willemsen, 2015, p.324). To detect a small effect (0.3 SD) with a power of 90% in a between-subjects study, a minimum of 150 participants per experimental condition is required (Knijnenburg and Willemsen, 2015, p.324). Eventually, 235 participants interacted with the original Keuzehulp and 199 users with the newly developed Keuzehulp. In **Table 1**, we have visualized the distribution of the participants based on their maximization tendency and level of indecisiveness. There can be seen that there is an even distribution of users. Since all participants were part of the Consumentenbonds' consumer panel and the data was anonymous, no informed consent was needed. The participants were informed that the data would (anonymously) be stored at the Eindhoven University of Technology (TUe).

Table 1. Distribution of participants with certain personal characteristics

Original Keuzehulp		
Level of indecisiveness	Maximization tendency	
	Low	High
Low	64	51
High	55	65

New Keuzehulp		
Level of indecisiveness	Maximization tendency	
	Low	High
Low	47	51
High	55	46

3.4. Measures

To test the hypotheses, we administer short questionnaires to measure participants' *Experience* (i.e., "Choice satisfaction" and "Perceived usefulness"), *Perceptions* (i.e., "Perceived guidance"), and *Personal Characteristics* (i.e., maximization tendency and level of indecisiveness). Most items used in this study were adapted from existing questionnaires. If necessary, slight adjustments were made to the questions, making them more in line with the goals of the current study. The complete questionnaire (Dutch and English) can be found in Appendix 1 (**Table 5**). For all items, a seven-point Likert scale was used, ranging from 1 ("Strongly disagree") to 7 ("Strongly agree"). The items of both the pre-and post-experimental questionnaire were submitted to Confirmatory factor analysis (CFA, as suggested by Knijnenburg and Willemsen 2015) to validate these scales. Factor loadings of items that were included are shown in **Table 3** and **Table 4**, as well as the average variance extracted (AVE) and Cronbach's alpha for each factor. For a complete summary of the items and the reliability scores of all the constructs, refer to the results section. In the following sections, the separate constructs will be discussed.

3.4.1. Maximization scale (PC)

This personal characteristic is used to distinguish individuals with respect to their maximization tendency (Nenkov et al., 2008). A distinction will be made between satisficers (low maximization) and maximizers (high maximization). The original Maximization scale developed by Schwartz et al. (2002) was comprised out of 13 items. As discussed in the introduction, one of the constructs was related to indecisiveness which is a construct on its own and should not be included in this scale. Dalal et al. (2015) shared these same beliefs. Therefore, we used their 7-Item Maximizing Tendency Scale (e.g., *No matter what I do, I have the highest standards for myself.* , Cronbach's Alpha =.86).

3.4.2. Indecisiveness (PC)

This personal characteristic was defined as a tendency to experience problems when making decisions across situations and domains (Crites, 1969). A distinction will be made between decisive people (low indecisiveness) and indecisive people (high indecisiveness). The most prominent measure of indecisiveness to date is called the Indecisiveness Scale (IS) and was created by Frost and Shows (1993). As discussed in the introduction, this scale has a couple of context-dependent items. In the current study, indecisiveness was measured using the six items with the highest factor loading in the study by Rassin et al. (2007) (e.g., *I like the product I've chosen*, Cronbach's Alpha =.79).

3.4.3. Perceived guidance (SSA)

This subjective system aspect denotes the extent to how well a system guides consumers in making trade-offs. This dependent variable was labeled as "Perceived guidance". These questions were tested statistically to establish content validity. Perceived guidance was measured through four items and was not based on an existing scale (e.g., *The Keuzehulp motivated me to make better trade-offs between products*, Cronbach's Alpha =.89).

3.4.4. Choice satisfaction (EXP)

Choice satisfaction was measured using four items (e.g., *I like the product I have chosen*, Cronbach's Alpha =.89) and was adapted from Knijnenburg, Willemsen, and Broeders (2014). This construct denotes the experience that the user has with the Keuzehulp.

3.4.5. Perceived usefulness (EXP)

Perceived usefulness denotes the degree to which a person believes that using a particular system would enhance performance (Venkatesh, Morris, Davis, and Davis, 2003). This construct was measured by using seven items (e.g., *Using the system would enable me to accomplish tasks more quickly*, Cronbach's Alpha =.92) and was adjusted from Knijnenburg, Reijmer, and Willemsen (2011) and Knijnenburg et al. (2014).

3.4.6. Behavioral measures

Participants' behavior was obtained by tracking the users' time spent on the simulation and how many critiques participants chose. Unfortunately, this data was not saved adequately for 172 participants, and therefore, in the remainder of this study, we only focused on the questionnaire data.

3.5. System design and implementation

3.5.1. User interface design

Both simulations were developed using Python (3.8). Only activities that were necessary for using the decision-making tool were made interactive. This ensured that participants would not get distracted by other parts of the website. Both simulations were developed for the category; washing machines. The original and newly developed Keuzehulp were both simulated using products that were available on the website in January 2021.

3.5.1.1. Original Keuzehulp

Fig. 3 (p. 8) illustrates the workflow of the original Keuzehulp. At the beginning of the interaction, participants answered four questions regarding attributes and product specifications. The decision tree with the questions can be found in Appendix 2 (**Fig. 13**). In addition, an example of a dialog window can be found in Appendix 3 (**Fig. 15**). Afterward, participants were presented a list of washing machines. This consideration set was generated using EBA, as explained in the introduction (Bettman, Luce, and Payne, 1998). Participants were able to change their answers at any time during the interaction.

Furthermore, participants could click on the individual products to look at the product page, providing more detailed information about the products. The simulation would end when a participant chose a product by pressing the button "Ja, dit wil ik" (**Fig. 7**). A pop-up screen would appear, which would direct them to the final questionnaire.

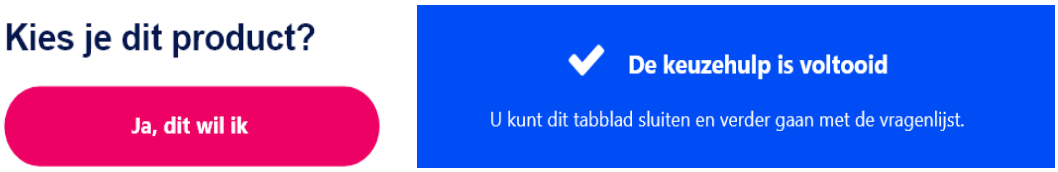

Fig. 7 Button to end simulation and notification

3.5.1.2. New Keuzehulp

Fig. 8 illustrates the workflow of the new Keuzehulp. The design of the new interface was kept as similar to the existing interface as possible. The workflow consists out of six steps: (1) A user answers two needs-based questions. (2) The system converts the answers into a set of filter settings. (3) The system generates three recommendations. Users either choose step 4 or step 6. (4) The user chooses to critique a product. (5) A new set of recommendations is generated. The user can again choose between step 4 of 6. (6) User chooses a product, and the simulation ends.

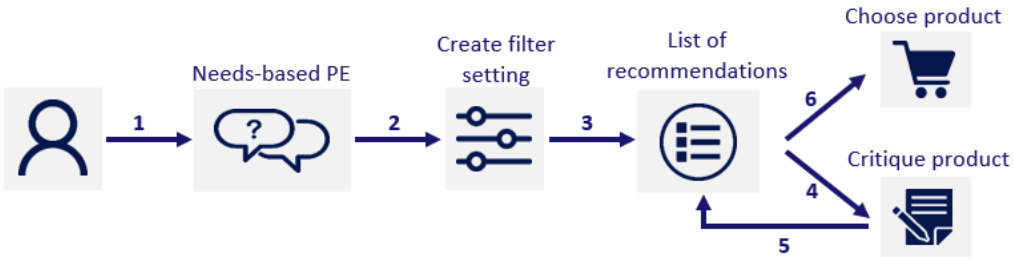

Fig. 8 Workflow newly developed Keuzehulp

To create the initial consideration set, a few questions were asked. These questions were chosen based on questions asked when using the original Keuzehulp and through consultation with a domain expert (Irene la Faille, Expert Woning, and Huishouden Consumentenbond,

December 2020). An example of a dialog window can be found in Appendix 3 (**Fig. 16**). The two questions included in the first part of the simulation were; "For how many people do you do laundry" and "What is your budget?". The decision tree with the questions and options to answer from can be found in Appendix 2 (**Fig. 14**). There are two reasons for asking these questions at the beginning of the interaction: 1) An initial consideration set has to be created for the system to recommend a set of products that participants can critique. 2) These two specifications were challenging to include in the critiquing part. In addition, the first question has nothing to do with constructing preferences; this is a requirement that does not change. A set of pre-defined set of attributes, identified by the Consumentenbond, was used (**Table 1**). This will be further elaborated on in the next section. After answering the two questions, participants were given a consideration set of three washing machines generated using EBA (Bettman, Luce, and Payne, 1998). Participants could click on the individual products that would direct them to a webpage, as shown in **Fig. 9**.

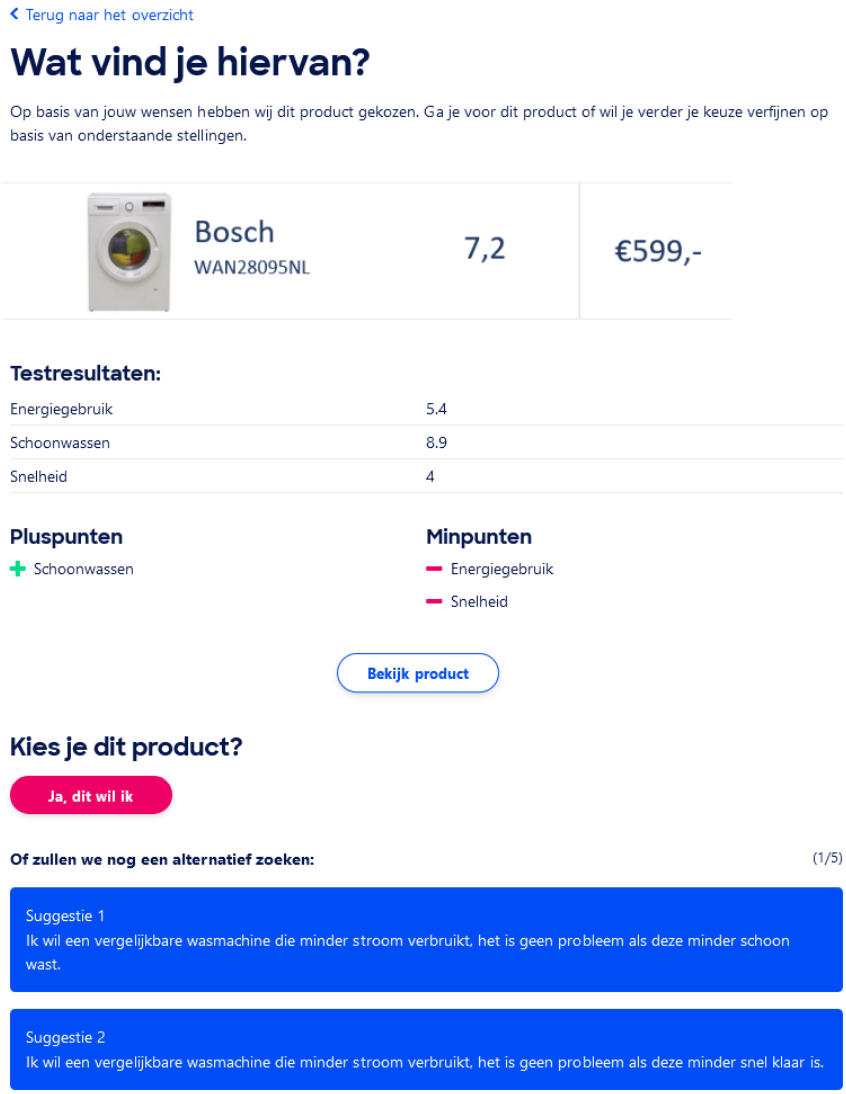

Fig. 9 Example critiquing page

Fig. 9 illustrates the layout of the critiquing page. On this webpage, users could find relevant product information about the selected product. The top part of the page was kept the same as the original Keuzehulp. The two blue buttons at the bottom of the webpage illustrate the two system-suggested critiques. When chosen, they returned a new set of recommendations. An example critique could be; "I would like a similar washing machine but one that rinses the clothes better, and it is no problem if it is less silent.". The attributes that could be addressed in the critiques were; "Cleanness", "Rinse", "Centrifugation", "Eco-friendly", "Water usage", "Speed", "Ease of use" and "Noise". In addition, information about the attributes addressed in the critiques is also shown on the web page. The participants could see what the selected product scored on these attributes and if these scores were considered below average or above average. These thresholds were chosen by a domain expert (Irene la Faille, Expert Woning, and Huishouden Consumentenbond, December 2020). The simulation would end when a participant chose a product and clicked on the button "Ja, dit wil ik" (**Fig. 7**, p. 25). A pop-up screen would appear, which would guide them to the final questionnaire.

3.5.2. Algorithm

In this section, the development of the new Keuzehulp will be discussed. As mentioned in the introduction, the Keuzehulp consists of two parts: 1) Generation of the initial consideration set. 2) Critiques are presented to the user. The non-compensatory strategy, EBA, was used to create the initial consideration set, and the compensatory strategy, preference-based critiquing, was used to generate the critique. Using these two steps allows consumers to first efficiently reduce a large set of alternatives to a more reasonable set and then thoroughly let them compare the products left to make the optimal decision. The generation of critiques was adopted from a heuristic method that was validated by Chen and Pu (2007c).

3.5.2.1. Recommendation Algorithm

First, an initial consideration set was created by asking the user a set of questions related to specific product specifications (Appendix 2, **Fig. 14**). Next, the answers were translated into filter settings to generate a consideration set of three washing machines. The three products were ranked based on their overall test scores, which were calculated by the Consumentenbond using MAUT. An example conversation can be found in Appendix 4 (**Fig. 17**). The user could either choose to pick one of these three products or critique one of the products.

3.5.2.2. Critiquing algorithm

Participants were given two system-suggested critiques for each product (e.g., "*I would like a similar washing machine but that rinses the clothes better and it is no problem if it is less silent*"). An

example of such a critiquing page can be found in **Fig. 9**. The personalized critiques were based on eight needs-based attributes (**Table 2**). Experts from the Consumentenbond identified these needs-based attributes and assigned importance weights to each of the eight attributes.

Table 2 Attributes and their importance weights

Product attributes	**Importance weights**
Cleanness ("Schoonwassen")	45%
Rinsing ("Spoelen")	10%
Centrifugation ("Centrifugeren")	10%
Eco-friendly ("Energieverbruik")	12%
Water usage ("Waterverbruik")	3%
Speed ("Snelheid")	5%
Ease of use ("Gebruiksgemak")	10%
Noise ("Geluid")	5%

The critiquing algorithm consisted out of four steps. First, all the products were converted into a trade-off vector, comprising of a set of (*attribute, trade-off*) pairs. This trade-off vector indicated whether a particular attributes' test score was *improved* ("Centrifugation, higher") or *compromised* ("Rinsing, lower") compared to the product that was being critiqued. If a product scored higher on a particular attribute, the trade-off vector would be ("Centrifugation, higher"). If the product scored lower on the attribute, the trade-off vector would be ("Centrifugation, lower"). An example trade-off vector could be: {(" Cleanness, equal"), ("Rinsing, lower"), ("Centrifugation, higher"), ("Eco-friendly, lower"), ("Water usage, lower"), (Speed, higher"), ("Ease of use, higher"), ("Noise, lower")}.

Next, to identify relevant critiques, an Apriori algorithm called association rule mining was used to discover the recurring and representative subsets of the attributes', trade-off pairs (Agrawal, R., Imielinski, T., Swami, A., 1993). The reason for applying this Apriori algorithm is due to its popularity and efficiency in mining associate rules among features (Chen and Pu, 2006).

3.5.2.3. Association rule mining

Association rules attempt to find groups of items that are commonly found together (Lai and Cerpa, 2001). Association rules are "if-then" statements that show the probability of relationships between items within large datasets. An association rule consists out of an antecedent and a consequent. The antecedents (if) is an item found within a dataset. The consequent (then) is an item found in combination with the antecedent. Essentially, the association is about discovering a

set of rules shared by a significant percentage of the data (Zaki, 2000). The rules used in this study consisted of two attributes since using more attributes could result in information overload (Le Lec, Lumeau, and Tarroux, 2016).

After applying the Apriori algorithm, all products were placed into categories. Where each category represented a critique (e.g., "rinses better", "less eco-friendly"). Products could belong to more than one category because they could have had different subsets of trade-off properties shared by other groups of products. Therefore, a large number of critiques can be generated. To be able to present the critiques that are most relevant for a consumer, critiques were ranked according to their trade-off utilities. In other words, looking at their gains and losses relative to the top candidate. By using this method, users' preferences were taken into account. This is the part of the algorithm that helps present the most relevant trade-offs to a consumer and helps figure out what attributes are most important to a person and which ones are not.

Next, the proposed critiques and their products were diversified to increase their suggestion power since similar items have limited added value (McGinty and Smyth, 2003). Diversification has been found to help the choice process (Konstan & Riedl, 2012; Knijnenburg et al., 2012) and increase choice satisfaction (Ekstrand et al., 2014; Langner & Krengel, 2013). An in-depth explanation of the preference-based algorithms can be found in Appendix 4.

After generating two relevant critiques, users were able to choose between the two proposed critiques. If participants chose one of the two critiques, a new consideration set would be suggested. This new consideration set consisted of three new products, which were chosen based on their average test score, which was calculated based on MAUT, as discussed earlier.

3.5.2.4. Personalization weights

The relative importance weights used to calculate the average test scores were refined after every critiquing cycle, personalizing the critiques. During the first cycle, the importance weights were based on the weights assigned by the Consumentenbond (**Table 2**, p. 28). During the subsequent critiquing cycles, the importance weights were updated based on users' critiquing history. For example, when a participant chose the critique, "rinses better", "less eco-friendly". The weight for rinsing would increase, and the weight for the attribute eco-friendly would decrease. This part of the algorithm makes our Keuzehulp personalized, and as a result, the critiques become personally relevant to the user. Since we start with the weights that were chosen by the Consumentenbond, the weights will become more personalized when consumers critique more products. Considering that we created simulations for our experiment, users could critique products up to a maximum of five times.

Chapter 4.: Results

In the following chapter, we examine the results of the experiment. The data were analyzed using the statistical software R (version 3.6.3). Firstly, the identified outliers will be described. Secondly, the construct validity of the questionnaires will be elaborated, followed by the analysis of the Structural Equation Model (SEM model).

4.1. Outliers

Data related to the duration of completing the questionnaire and simulation were inspected, and the answers to the questions. Before checking for outliers, a log-transformation was conducted on the time variable (duration of the experiment and duration of using the simulation) to transform the skewed data to approximately conform to normality. However, the data still did not meet the normality assumption for both the duration of the simulation and the duration of the questionnaire. To detect if there were any outliers in the data, the Z-score method was used. Since both variables did not meet the normality assumption, a $|z| > 4$ was used as a threshold to detect outliers. A total of three outliers were discovered (participants: 296, 260, and 390). After visually inspecting the data, two other participants were also considered as outliers since they spent over five hours on the questionnaires (Participants: 132 and 408). No outliers or missing values were detected in the answers to the questionnaire. The statistical tests will be performed twice, once with outliers and once without. Depending on whether the outliers substantially influence the outcomes and the interpretation of the results, there will be decided to either include or exclude the considered outliers.

4.2. Establishing construct validity

First, construct validity must be established to ensure that the items comprise accurate and robust measurement scales. We performed a Confirmatory Factor Analysis (CFA, as suggested by Knijnenburg and Willemsen 2015) on all the items of both the pre-and post-experimental questionnaires. Since a 7-point Likert scale was used to answer the questions, the CFA used ordinal dependent variables. Firstly, convergent and discriminant validity will be tested. Items with low factor loadings, high residual correlations, or high cross-loadings were removed from the model. The factor loadings of the construct's items are shown in **Table 3** and **Table 4**, as well as the average variance extracted (AVE) and Cronbach's alpha. The Cronbach's alpha and the values of AVE are used to check for convergent validity. When the square root of the AVE's is higher than any of the factor correlations, discriminant validity is held.

4.3. Pre-experimental questionnaires

4.3.1. Maximization tendency

Inspecting the maximization scale, all the items show good factor loadings (**Table 3**). The values of the AVE and Cronbach's alpha are adequate, indicating convergent validity. The square root of the AVE's is higher than any of the factor correlations, indicating that discriminant validity is also accepted.

4.3.2. Level of indecisiveness

Though we selected the six items from an original scale (of 11 items), it seems that some items did not fit well. Convergent validity does not hold with an AVE of 0.48. Therefore, we decided to exclude the items that reduced the fit, to improve the scale. In the final model, items two, three, and six were removed since they had low factor loadings (**Table 3**). After removing the three items, the AVE was 0.60, indicating that convergent validity is accepted. The square root of the AVE's is also higher than any of the factor correlations, indicating that discriminant validity is achieved.

Next, the covariances were examined. The correlation between maximization and indecisiveness is low and non-significant, with an estimate of -0.07 (SE = 0.05, p = 0.14) (**Appendix 5**: Statistical output, p. 67). Indicated that the two constructs are indeed different. This is in line with research by Cheek and Goebel (2020), who argued that these two constructs measure different things and subsequently should be measured separately.

Table 3 Items presented in the pre-questionnaires. Items without a factor loading were excluded from the analysis.

Construct	Items	Factor loading
Maximization	**1. Wat ik ook doe, ik stel altijd hoge eisen.**	**0.602**
Alpha: 0.86	**2. Ik neem nooit genoegen met het op één na beste product.**	**0.625**
AVE: 0.55	**3. Ik doe er altijd alles aan om het beste product te vinden.**	**0.505**
	4. Goed genoeg is voor mij niet voldoende.	**0.554**
	5. Ik probeer altijd het maximale resultaat te halen.	**0.511**
	6. Ik wacht tot ik het beste product vind, ongeacht hoe lang het duurt.	**0.500**
	7. Ik sluit nooit een compromis wanneer ik een product wil aanschaffen.	**0.541**
Indecisiveness	**1. Ik vind het makkelijk om keuzes te maken.**	**0.700**
Alpha: 0.79	**2. Na het maken van een keuze, twijfel ik of ik wel de juiste beslissing heb genomen.**	
AVE: 0.60	**3. Wanneer ik een keuze heb gemaakt, ben ik vrij zeker dat het de juiste keuze is.**	
	4. Ik stel het maken van keuzes vaak uit.	**0.419**
	5. Meestal maak ik snel keuzes.	**0.689**
	6. Ik vind het fijn om degene te zijn die verantwoordelijk is voor het maken van keuzes.	

4.4. Post-experimental questionnaires

4.4.1. Choice satisfaction

Looking at the construct choice satisfaction, item three has a low factor loading and was therefore removed (**Table 4**). AVE and Cronbach's alpha are good, indicating convergent validity. The square root of the AVE's is higher than any of the factor correlations, indicating that discriminant validity is also accepted.

4.4.2. Perceived usefulness

Inspecting the measure perceived usefulness, item one shows a low factor loading (**Table 4**). While inspecting the modification indices, we found that this item also loaded strongly on the latent variable choice satisfaction and was therefore removed due to cross-loadings (**Appendix 5**, p. 67). Looking at the modification Indices, we also found that items six and seven were highly correlated, and therefore item seven was removed, improving the model fit. Item seven was removed rather than item six since this item had a lower factor loading compared to item six. Convergent validity holds with an AVE of 0.72.

4.4.3. Perceived guidance

All factor loadings of perceived guidance are higher than .5, and the AVE was 0.75, indicating convergent validity (**Table 4**). Looking at the covariances between perceived guidance and the other constructs, we observe that perceived guidance correlates highly with perceived usefulness. The R^2 is higher than the square root of the AVE's (R^2= 0.82), indicating that discriminant validity might not hold.

Although we started with five constructs, there were some cross-loadings between the constructs' items, implying that some items are not loading on their construct or are loading on other constructs (**Appendix 5**, p. 67). Together with insights from our Confirmatory Factor Analysis, discriminant validity does not seem to hold for the two constructs; perceived usefulness and perceived guidance. Therefore, we ran a preliminary analysis of the hypothesized SEM model in three variants. Once the variable perceived guidance was excluded from the model, once the variable perceived usefulness was excluded, and once the two variables were combined into a single variable. Comparing these models, we found that excluding the construct, perceived usefulness resulted in a better-fitting model compared to the other two, and all R^2 values were acceptable. Subsequently, improving users' perceived guidance was our main goal in improving the Keuzehulp and answering the two research questions. By removing perceived usefulness, the

original SEM model will change, and subsequently, some of our hypotheses will no longer be tested.

Table 4 *Items presented in the post-questionnaires. Items without a factor loading were excluded from the analysis*

Construct	Items	Factor loading
Choice satisfaction Alpha: 0.89 AVE: 0.77	**1. Ik ben tevreden met de wasmachine die ik gekozen heb.**	**0.862**
	2. Ik vind dat ik de beste wasmachine heb gekozen uit de lijst met opties.	**0.864**
	3. De door mij gekozen wasmachine past precies bij mijn voorkeuren.	
	4. Ik had liever een andere wasmachine gekozen.	**0.580**
Perceived usefulness Alpha: 0.92 AVE:0.81	**1. Ik kon maar een beperkt aantal voorkeuren aangeven in de Keuzehulp.**	
	2. Ik wil deze Keuzehulp vaker gebruiken, ook bij andere producten.	**0.855**
	3. Met behulp van de Keuzehulp kan ik betere keuzes maken.	**0.839**
	4. De Keuzehulp was nutteloos.	**0.709**
	5. Ik zou de Keuzehulp aan anderen aanraden.	**0.880**
	6. De Keuzehulp zorgt ervoor dat ik sneller keuzes kan maken tussen verschillende wasmachines.	**0.753**
	7. De Keuzehulp zorgt ervoor dat ik op eenvoudige wijze een wasmachine kan kiezen.	
Perceived guidance Alpha: 0.89 AVE: 0.75	**1. De Keuzehulp maakt het makkelijker voor mij om te kiezen welke eigenschappen van de wasmachines ik belangrijk vind.**	**0.808**
	2. De Keuzehulp motiveert me om betere afwegingen te maken tussen de wasmachines.	**0.698**
	3. De Keuzehulp is een eenvoudige manier om de best passende wasmachine voor mij te vinden.	**0.776**
	4. De Keuzehulp helpt me bij het begrijpen van de belangrijke eigenschappen van de wasmachines.	**0.711**

4.5. Structural model

To investigate the research questions and their subsequent hypotheses, the data were analyzed using structural equation modeling (SEM, as suggested by Knijnenburg and Willemsen 2015). A path model was created using the Lavaan package in R. For the analysis, the recommendations by Knijnenburg and Willemsen (2015) were followed, starting with a saturated model and consecutively pruning non-significant effects from the model. In order to calculate the moderations, we used the factor scores from the variable maximization tendency and level of indecisiveness rather than adding these variables as latent constructs to the SEM model. Next, the center mean was calculated for the variables' maximization tendency, level of indecisiveness, and the dichotomous variable, Keuzehulp. In the end, the outliers did not affect the outcome of the results. Therefore, the outliers were not removed from the data.

The final model had a good model fit ($\chi^2(40, N=343) = 88.11$, $p < 0.001$, CFI = 0.994, TLI = 0.997 and RMSEA= 0.05, 90% CL[0.038, 0.068]) (**Appendix 5**, p. 66). **Fig. 10** displays the effects

that were found in this model, showing the standardized path coefficients. The direction of the arrows demonstrates the nature of the relationship. The numbers on the arrows denote the regression coefficients with the standard errors displayed in the parentheses. Subsequently, the p-values below these two numbers denote the statistical significance of the effect.

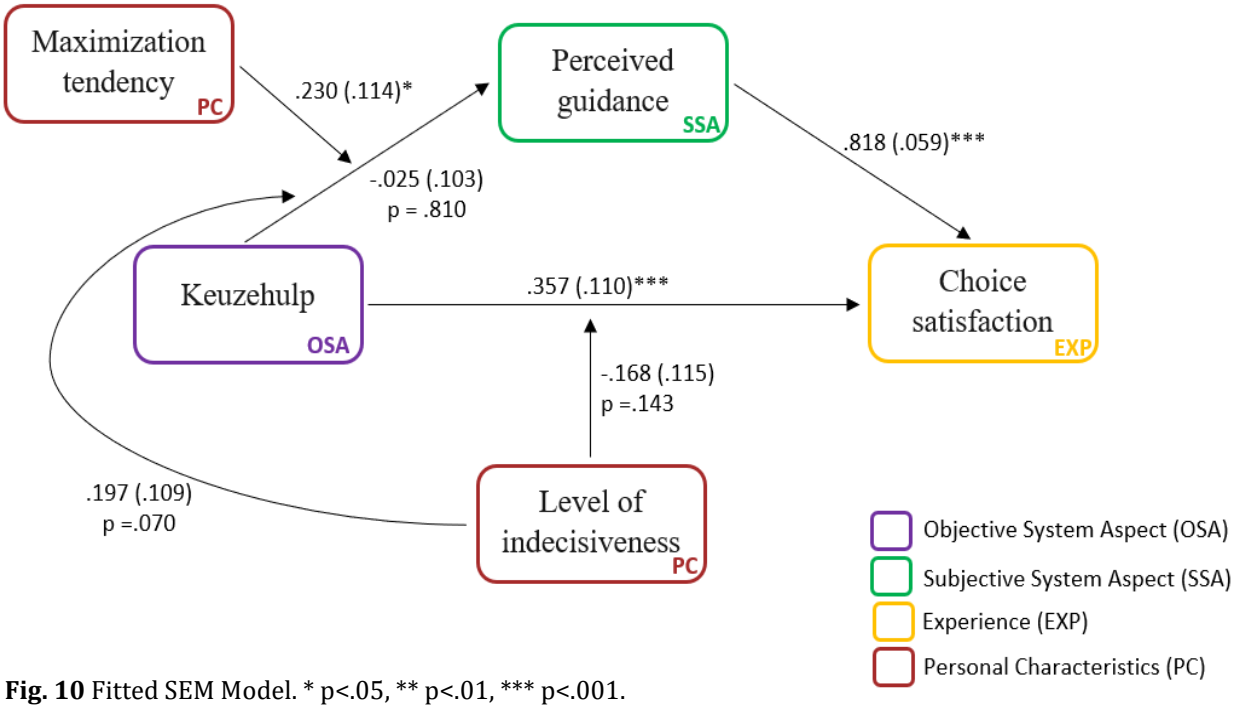

Fig. 10 Fitted SEM Model. * p<.05, ** p<.01, *** p<.001.

In the final model, maximization tendency did not relate to choice satisfaction and was removed from this path. Subsequently, the interaction variable, Keuzehulp*Maximization, was removed from this path. The other variables were either significant or, when removed, would make other correlations non-significant. Next, the individual hypothesis will be discussed.

The results show that no direct effect of the Keuzehulp on perceived guidance was found (p = .810). Meaning that there is no significant difference between the two decision-making tools and the participants' score of perceived guidance (H1 not supported). In terms of choice satisfaction, we observe that this is significantly (and positively) influenced by the new Keuzehulp (H2 is supported). Notably, participants who used the new Keuzehulp had on average higher self-reported choice satisfaction (M = 0.08, SE = 0.07), compared to the participants that used the original Keuzehulp (M = -0.09, SE = 0.06). Next, participants reported higher choice satisfaction when they perceived more guidance, as shown by the direct effect of perceived guidance on choice satisfaction (p < .001) (H3 is supported). Since we did not find a significant direct effect of the Keuzehulp on perceived guidance, no mediation effect was found of perceived guidance on Keuzehulp and choice satisfaction (H4 not supported). Meaning that perceived guidance does not directly explain the effect of the Keuzehulp on choice satisfaction.

To test hypotheses H6-H9, moderation analysis was conducted, with the two personal characteristics as the moderators; maximization tendency and level of indecisiveness (**Fig. 10**).

4.5.1. Perceived guidance

First, we looked at the interaction effect of the two personal characteristics and the Keuzehulp on perceived guidance. To better understand these interaction effects, we plotted the mean perceived guidance score for people who used the original Keuzehulp and participants who used the new Keuzehulp, as presented in **Fig. 11**. We started by creating dichotomous variables for maximization and indecisiveness (1: high score and 0: low score) using a median split. The grey line in **Fig. 11**. represents the group of participants with low maximization (satisficers), and the orange line represents the effect for people in the high maximization group (maximizers). Additionally, the mean perceived guidance score of satisficers and maximizers can be found in **Appendix 5** (p. 65).

In terms of the interaction effect of Keuzehulp*maximization on perceived guidance (**Fig. 10**), we observe that this effect is significant, with an estimate of 0.230 (SE = 0.114, p = 0.04) (H6 is supported), which means that users' maximization tendency moderates the effect of Keuzehulp and perceived guidance. This effect can also be observed in **Fig. 11**. Notably, we can see that maximizers who used the new Keuzehulp perceived more guidance, and satisficers perceive less guidance (**Fig. 11**). No difference was observed between satisficers and maximizers who used the original Keuzehulp.

Looking at the level of indecisiveness (**Fig. 10**), we found no significant interaction effect of Keuzehulp*indecisiveness on perceived guidance (p = 0.07) (H7 is not supported). However, inspecting the effects of the decisive and indecisive people as dichotomous variables, we observe that indecisive participants who used the new Keuzehulp perceived more guidance compared to decisive individuals (**Fig. 11**). No significant difference in perceived guidance was found between decisive and indecisive participants who used the original Keuzehulp.

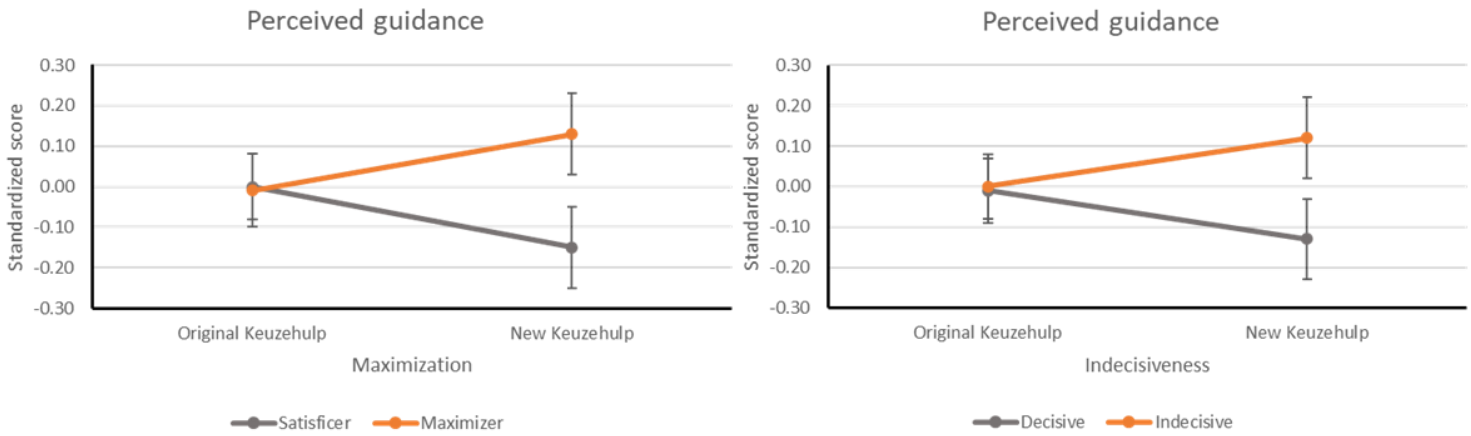

Fig. 11 Mean perceived guidance for satisficers and maximizers and decisive and indecisive users for the original Keuzehulp and New Keuzehulp. Error bars are one standard error of the mean.

4.5.2. Choice satisfaction

Next, we looked at the interaction effect of the two personal characteristics and the Keuzehulp on choice satisfaction. To better understand this interaction effect, we plotted the mean choice satisfaction scores for people who used the original Keuzehulp and participants who used the new Keuzehulp, as presented in **Fig. 12**. The grey line in **Fig. 12** represents the group of participants with low maximization (satisficers), and the orange line represents the effect for people in the high maximization group (maximizers). The mean choice satisfaction of satisficers or maximizers can be found in **Appendix 5** (p. 65).

Looking at our SEM model (**Fig. 10**), we did not find an interaction effect between users' maximization tendency and Keuzehulp on choice satisfaction (H8 is not supported). Meaning that users' maximization tendency does not moderate the effect of Keuzehulp and choice satisfaction directly. However, looking at the effects of users' maximization tendency as a dichotomous variable, we can see that users' maximization tendency might moderate, suggesting that other variables might influence this moderation (**Fig. 12**). Notably, we can see that maximizers who used the new Keuzehulp were happier with their choice compared to satisficers.

Finally, we investigated the interaction effect of Keuzehulp and users' level of indecisiveness on choice satisfaction. Our SEM model shows no significant interaction effect of Keuzehulp*indecisiveness on choice satisfaction (**Fig. 10**) (H9 is not supported), which means that users' level of indecisiveness does not moderate the effect of Keuzehulp and choice satisfaction directly. However, looking at the effects of decisive and indecisive people as dichotomous variables, we can see those indecisive people who used the new Keuzehulp were happier with their choice than decisive participants who used the same Keuzehulp (**Fig. 12**). A similar effect was observed for participants who used the original Keuzehulp; indecisive participants had more choice satisfaction when compared to decisive participants.

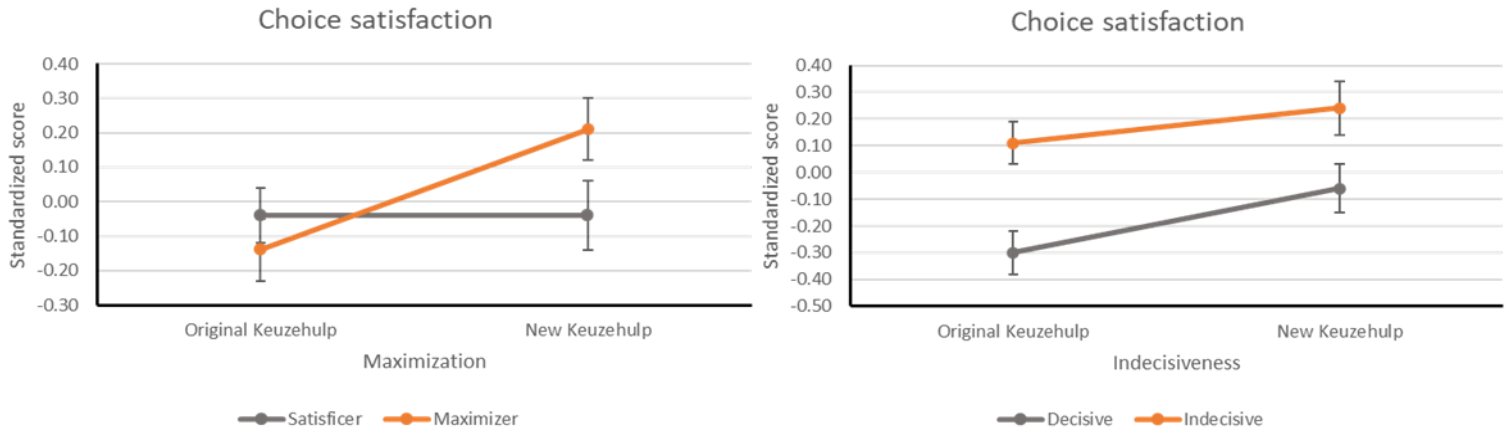

Fig. 12 Mean choice satisfaction for satisficers and maximizers and decisive and indecisive users for the original Keuzehulp and New Keuzehulp. Error bars are one standard error of the mean.

4.5.3. Moderated mediation

As discussed earlier, there is a high positive correlation between perceived guidance and choice satisfaction. Due to this high correlation, the interaction effect of Keuzehulp*maximization on choice satisfaction might come from perceived guidance. Meaning that maximizers have higher choice satisfaction because they perceived more guidance, suggesting that there might be a moderated mediation. We reran the SEM model with the moderation and mediation embedded in the model (**Appendix 5**, p. 65). We observe a statistically significant mediation of Keuzehulp on choice satisfaction through perceived guidance, with an estimate of 0.187 (SE = 0.09, p = 0.04). Implying that when someone is a maximizer and uses the new Keuzehulp, they perceive more guidance, which subsequently increases their choice satisfaction.

We performed the same analysis for the interaction effect of Keuzehulp*indecisiveness (**Appendix 5**, p. 66). The moderated mediation of Keuzehulp and indecisiveness on choice satisfaction through perceived guidance was almost significant, with an estimate of 0.161 (SE = 0.09, p = 0.07). Suggesting that indecisive users who use the new Keuzehulp are more satisfied with their choice because they perceived more guidance. However, we need to take into account that this effect is non-significant.

Chapter 5.: Discussion

The Consumentenbond has developed a decision support tool to help consumers in finding a product that is in line with their needs and wishes. Like most online (shopping) environments, they use a non-compensatory strategy to help consumers in finding a product, which has been shown to lead to sub-optimal results since trade-offs are being ignored. Considering there are well-supported findings that compensatory-based support is normatively better, our goal was to improve their decision support tool by using compensatory and personalized support. Furthermore, researchers have argued that not everyone might appreciate the same type of support. Therefore, we investigated two personal characteristics which might influence the perception and experience of the new tool.

Critiquing has been found to guide consumers in efficiently finding their ideal product, which is particularly meaningful when consumers are searching for durable products (Chen and Pu, 2009). To investigate if adding a critiquing-based element to the Keuzehulp would increase consumers' perceived guidance in making trade-offs and improve their choice satisfaction, a new Keuzehulp was developed. Furthermore, we looked if users' maximization tendency and their level of indecisiveness influence their perception and experience with the Keuzehulp. The new Keuzehulp utilizes a non-compensatory and compensatory strategy. The non-compensatory strategy, EBA, was used to recommend an initial consideration set. The compensatory strategy, Preference-based critiquing, was used to guide consumers in making the necessary trade-offs. A series of subjective constructs have been tested with questionnaires to gain insights into elements such as choice satisfaction and perceived guidance.

Next, the main findings of the study are summarized, and the study's limitations are assessed. Finally, an overall conclusion is drafted together with some additional suggestions for possible future research.

5.1. Main Findings

This study aimed to determine if developing a new decision support tool could help consumers make normatively better decisions. We tested nine hypotheses, each of which will be discussed below.

First, we investigated if user's who used the new Keuzehulp would perceive more guidance than people who used the original Keuzehulp (H1). We expected that this would indeed be the case since researchers have argued that critiques carry considerable explanatory power because they show users the trade-offs that have to be made (McCarthy, Reilly, McGinty, and Smyth, 2004). Furthermore, researchers have argued that critiquing recommender systems may guide consumers in efficiently finding their ideal product (Chen and Pu, 2009). The results show no significant direct difference in perceived guidance between the original Keuzehulp and the new Keuzehulp. One possible explanation for this finding is that we have created a simulation for the new Keuzehulp. Due to time constraints, there were several design features that we were not able to implement in the simulation. For example, we could not provide users with information about the attributes (e.g., what they mean). This could be a possible reason why the new Keuzehulp did not optimally guide people in choosing between trade-offs since the attributes might not have been clear enough. However, we found that this effect was moderated by other variables. This will be further elaborated on later in the discussion.

Next, our results shed some light on the effect of the Keuzehulp on participants' choice satisfaction. As the choice process becomes more personalized, we would expect people to be happier with the choice they make. Our results support these expectations, suggesting that personalizing the Keuzehulp indeed increases users' choice satisfaction (H2). This is in line with earlier research that providing personalized recommendations based on user's preferences improves the accuracy of a system to predict the best recommendations for the user (Burke, 2002). This subsequently results in a better user experience regarding choice satisfaction (Knijnenburg, Willemsen, Gantner, Soncu, H., and Newell, 2012). In addition, Chen and Pu (2007) argued that using a preference-based critiquing algorithm achieves significantly higher critique prediction accuracy compared to other systems that only used purely data-driven algorithms. Another explanation for these findings is that a combination of non-compensatory and compensatory strategies was used to develop the new Keuzehulp. Combining the two decision-making strategies allows people to efficiently reduce a large set of alternatives to a more reasonable set and then thoroughly compare the products that are left to make a more optimal decision (Xiao and Benbasat, 2007).

Looking at our model, we found a high correlation between perceived guidance and choice satisfaction (H3). People who perceive more guidance had significantly more choice satisfaction. A possible explanation for this high correlation could be that when people perceive more guidance, they will become more confident in the choices they make. Since we did not find a significant direct effect of the Keuzehulp on perceived guidance, no mediation effect was found of perceived guidance on Keuzehulp and choice satisfaction (H4). In other words, people were not

more satisfied with the new Keuzehulp than participants who used the original Keuzehulp because they perceived more guidance.

Several researchers have argued that situational or personal aspects could influence how people perceive and evaluate a system (e.g., domain knowledge; Knijnenburg and Willemsen, 2009, 2010; Knijnenburg et al., 2012; Kamis and Davern, 2004). Since we suspected that users' maximization tendency and level of indecisiveness might influence these variables, we tested these variables, which will be discussed next.

5.1.1. Maximization tendency

We investigated if maximizers perceived more guidance when using the new Keuzehulp compared to satisficers. The results of this study provide evidence that there is indeed a significant moderating effect of users' maximization tendency on the Keuzehulp and perceived guidance (H6). Maximizers significantly perceived more guidance compared to satisficers. One possible explanation could be that maximizers often tend to spend a lot of time and effort when comparing products and aspire to find the best possible option (Schwartz, 2004). Since the new Keuzehulp shows individuals the different trade-offs that have to be made, users might subsequently learn more about the products on the spot, which might help them justify their decisions and choose a product accordingly.

Furthermore, we investigated if there was a difference in choice satisfaction between maximizers and satisficers. We assumed that maximizers had more choice satisfaction when using the new Keuzehulp compared to satisficers (H7). The results show no significant (moderating) effect of maximization on Keuzehulp and choice satisfaction. Nevertheless, we did observe (Fig. 12) that users' maximization tendency affects their choice satisfaction. Maximizers were more satisfied with their choice when they used the new Keuzehulp compared to satisficers. A possible explanation might be that the original Keuzehulp is non-compensatory. As discussed in the introduction, using this type of strategy does not help people make trade-offs, which is an important aspect of high-quality rational decision-making. In addition, existing evidence strongly suggests that compensatory strategies provide better decision quality to users compared to non-compensatory strategies (Pereira, 2001). Maximizers usually engage in extensive product comparison and aspire to find the best possible option (Schwartz, 2004). Since the original Keuzehulp only asks four general questions, a maximizer might not feel confident enough when their decision is based on only four questions. Furthermore, maximizers might be more satisfied with their choice when using the new Keuzehulp because this system provides personalized recommendations based on user's preferences, which has been proven to increase the accuracy of a system to predict the best recommendations for the user (Burke, 2002). It may also be that any system that displays more details and has more explanatory power increases maximizers'

confidence in their choice and subsequently increases their choice satisfaction. Satisficers, on the other hand, were not satisfied with their choice when using the new Keuzehulp. It may be that satisficers prefer the original Keuzehulp because it is less complex. Studies have shown that satisficers quickly want to attain satisfactory results (Cheek and Goebel, 2020). When individuals do not aspire to find the best possible option, they might already have enough justification for their choice based on the four questions.

Due to the high correlation between perceived guidance and choice satisfaction, we expected that there might be a moderated mediation of users' maximization tendency. The moderated mediation analysis supports our expectations. Maximizers significantly perceived more guidance and, in turn, had more choice satisfaction. As discussed above, the new Keuzehulp shows individuals the different trade-offs that have to be made. Users might therefore learn more about the product on the spot, which might help them justify their decisions, which might make them more confident and satisfied with the choice they made.

5.1.2. Level of indecisiveness

Looking at users' level of indecisiveness, we investigated if there was a difference in perceived guidance between decisive and indecisive people (H8). This study shows that this moderation effect is not significant. As discussed earlier, we were not able to implement several design features due to time constraints. Since research has shown that indecisive individuals often use a lot of information to make a choice (Ferrari and Dovidio 2000; Rassin 2007), they might not have felt confident enough with their choice since they might not have understood all of the attributes.

Furthermore, we investigated if there was a difference in choice satisfaction between decisive and indecisive people (H9). We observed that decisive people had less choice satisfaction than indecisive users. A discussed before, indecisive people tend to show a desire to seek out more information. These individuals often use this information to confirm that they are making the right choice (Ferrari and Dovidio 2000; Rassin 2007). The original Keuzehulp consists of only four general questions, which might not convince an indecisive person that they have made the right choice.

As discussed above, a high correlation was found between perceived guidance and choice satisfaction. Therefore we expected that there might be a moderated mediation of users' level of indecisiveness. The moderated mediation analysis showed that there was almost significant. Suggesting that indecisive people perceived more guidance and subsequently have higher choice satisfaction. A similar explanation to maximization tendency could be given for this personal characteristic. The new Keuzehulp shows users the trade-offs that have to be made. Since indecisive people tend to desire to seek out a lot of information, this information might help them

justify their decision, which might result in more perceived guidance and, in turn, users become more satisfied with their choice. However, all effects of this personal characteristic were non-significant. Therefore we cannot make any hard claims about this variable. Additional research is be needed to investigate this further.

5.2. limitations

The present study was subject to several limitations, which will be described in the following section. Furthermore, we will discuss some implications we encountered during the study.

Firstly, both simulations were developed in collaboration with a developer from the Consumentenbond (D. de Zeeuw). Due to limited time, we were not able to implement all the aspects of the original Keuzehulp. The product pages were initially screenshots of the information that was available on the website of the Consumentenbond. Therefore many buttons were visible in the simulation but were inactive. The participants were informed about the interactivity of the simulation. However, it is still unclear what effects the UX design had on the participants.

Furthermore, we tried to keep the two simulations as similar as possible, but this was not always possible. For example, we only showed participants the top three products when they used the new Keuzehulp. Depending on the available products in the original Keuzehulp, more than three products were shown to the participants. Schwartz (2004) found that the amount of products shown to a user influences their perception and interaction with a system. Since we were not able to look at the participants' behavior when interacting with the simulation, we could not see how many products were shown to the participants.

Secondly, we did not ask the participants if they had ever used the original Keuzehulp before. Their experience or interpretation could differ depending on whether or not they were already familiar with using the original Keuzehulp.

Thirdly, we showed the users washing machines that were available in January. Due to some changes in policies, the number of available washing machines was limited. If more washing machines were available, the algorithm would have based the recommendations on more products. On a couple of occasions, only one or two products were recommended instead of the minimum of three. This might negatively affect the users' perception of the systems' capabilities.

Fourthly, most items used in this study were adapted from existing questionnaires. Since all participants were Dutch, we had to translate the English scales. Generally, translating a scale directly from one language to another does not guarantee the content equivalence of the

translated scale (Brislin, 1970; Sechrest and Fay, 1972). Furthermore, the scales that were used for indecisiveness and maximization were general tendency scales. It is not clear if people who are "in general" maximizers or indecisive have these tendencies when buying a product. Cheek and Goebel (2020) argued that individuals might differ in their need to maximize depending on the context. Schwart et al. (2002) observed that domain-specific maximization scales might predict domain-specific outcomes better than a general maximization scale. Regarding the indecisiveness scale, we only used a small part of the scale because we wanted to limit the number of questions participants had to answer.

Lastly, as this was a field experiment, the usual control of a laboratory study was lacking. We have to take into consideration that the participants were not looking to purchase a washing machine. When people are looking for a washing machine, they will probably spend more time and attention on finding the right product. In addition, we only tested the Keuzehulp for the product domain; washing machines. To investigate if the same effects are found in other product domain areas, more research is needed.

5.3. Implications

One of the implications we came across is that we tried to track participants' behavior when using the Keuzehulp. Data was only saved when participants pressed the button "Ja, dit wil ik" (Fig . 8, p.25). Unfortunately, a large part of the group did not press this button, even though it was clearly stated in the instruction. Therefore, we were not able to save their behavioral data.

Furthermore, our goal was to investigate users' perceived usefulness of the Keuzehulp. Unfortunately, this construct correlated too highly with choice satisfaction and was therefore excluded from our analysis. However, this is a very important variable that should be considered in the future since research has found that perceived usefulness is an important driver of consumers' online shopping intentions (Amin, 2007). It is therefore essential that in further research, perceived usefulness is also tested.

Considering the limitations and implications mentioned above, the possibilities for future research will be highlighted in the next chapter.

Chapter 6.: Conclusion and Future research

Concluding, this thesis shows a promising direction for the field of decision support tools research. As discussed in this thesis, most online (shopping) environments still use non-compensatory strategies to help guide their consumers. Even though there are well-supported findings that compensatory-based support is normatively better. The current study has further proven these findings. Therefore we encourage these online e-commerce websites to go beyond using non-compensatory strategies to support consumers. In addition, we found evidence that personalizing the decision support tool and guiding consumers in making trade-offs improved users' perception and experience. More specifically, this study has revealed that people are more satisfied with their choice when using the newly developed Keuzehulp. Our findings also suggest that users' perceived guidance is moderated by their maximization tendency. Subsequently, we found that maximizers had more perceived guidance when using the new Keuzehulp, which in turn resulted in more choice satisfaction. Suggesting that it would be beneficial to create a decision support tool that tailors to the users' maximization tendency (Randall, Terwiesch, and Ulrich, 2007).

6.1. Future research

Further research is needed to investigate the reproducibility of the Preference-based critiquing algorithm. Due to time constraints, the optimal value for altering the importance weights was not empirically validated. Further research could also look into consumers' critiquing behavior to investigate what kind of critiques people prefer and how often they would like to critique. Moreover, looking at the users' behavior would help better understand the users' interaction with the Keuzehulp. Knijnenburg and Willemsen (2015) discussed that subjective components like your maximization tendency might explain the observed behavior (i.e., number of critiquing cycles or time spent on finding a product). At the same time, behavior grounds the subjective part of the evaluation in observable behavior.

Secondly, we suggest that future research could look into other factors that might influence the perception and interaction with the Keuzehulp, like situational and other personal characteristics. In this study, we have looked into two personal characteristics that might affect the outcome. However, Willemsen, Graus, and Knijnenburg (2016) have pointed out that there are numerous personal (e.g., domain knowledge) and situational (e.g., choice goals) characteristics

that could influence the way people perceive and interact with a system. It would therefore be interesting to look into other characteristics.

Lastly, it would be interesting to look at the general tendency of members of the Consumentenbond. For example: Are most of their member maximizers or satisficers? What types of members choose to use the Keuzehulp? Moreover, we have conducted a quantitative study. However, a qualitative study could provide more insights in identifying the expectation and needs of consumers, for example, by showing the two decision-making tools to the participants and ask their opinions (using a within-subjects design), to get more insight into the design aspects and the usability of the Keuzehulp.

6.2. Recommendations for the Consumentenbond

Overall, this study has provided evidence that creating a tool that uses a compensatory strategy and takes users' preferences into account might improve the current Keuzehulp. However, since this is the first study looking into the new Keuzehulp, it is important to test this system further. Even though this study has shown that the Keuzehulp might improve user support, this type of support may take a lot of time and effort to implement. To implement this on the website for all products, the tool must be personalized for each product category since the algorithm uses product information to create recommendations. Therefore further research is needed not make sure that this tool will indeed improve the decision support.

Aside from the general suggestion for future research that has been discussed above, the Consumentenbond could look into creating user profiles. Up to this point, they have not been able to create these user profiles since their members often share their accounts with family members. To solve this problem, the Consumentenbond could create user profiles for each individual family member. For example, Netflix users can create separate accounts with one membership, making it easier to create individual user profiles. Having a user profile could give the Consumentbond more possibilities and different recommendation techniques could further be investigated (e.g. collaborative filtering).

References

Agrawal, R., Imieliński, T., and Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, 207-216.

Aksoy, L., Bloom, P. N., Lurie, N. H., and Cooil, B. (2006). Should recommendation agents think like people?. *Journal of Service Research*, *8*(4), 297-315.

Alan, A. K., Kabadayi, E. T., Bakis, S., and Ildokuz, S. I. (2017). The role of individual factors on online shopping behavior. *Journal of Business Economics and Finance*, *6*(3), 200-215.

Amin, H. (2007). Internet banking adoption among young intellectuals. *Journal of Internet Banking and Commerce, 12*(3), 1-13.

Babutsidze, Z. (2012). How do consumers make choices? A survey of evidence. *Journal of Economic Surveys*, *26*(4), 752-762.

Bettman, J.R. (1971) The structure of consumer choice processes. Journal of Marketing Research, 465–471.

Bettman, J. R. (1979). *Information processing theory of consumer choice*. Addison-Wesley Pub. Co.

Bettman, J., Luce, M., and Payne, J. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research, 25*(3), 187-217.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, *1*(3), 185-216.

Burke, R. (2000). Knowledge-based recommender systems. *Encyclopedia of library and information systems*, *69*, 175-186.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, *12*(4), 331-370.

Cheek, N. N., and Goebel, J. (2020). What does it mean to maximize? "Decision difficulty," indecisiveness, and the jingle-jangle fallacies in the measurement of maximizing. *Judgment and Decision Making*, *15*(1), 7.

Chen, C. W. D. and Cheng, C. Y. J. (2009). Understanding consumer intention in online shopping: a respecification and validation of the DeLone and McLean model. *Behavior and Information Technology, 28*(4), 335-345.

Chen, L., and Pu, P. (2004). Survey of Preference Elicitation Methods. *EPFL Technical Report IC*, 1–23.

Chen, L., and Pu, P. (2006). Evaluating critiquing-based recommender agents. In *AAAI*, *6*, 157-162).

Chen, L., and Pu, P. (2007). Hybrid critiquing-based recommender systems. *Proceedings of the 12th international conference on Intelligent user interfaces - IUI '07*, 22–31.

Chen, L., and Pu, P. (2007c). Preference-Based Organization Interfaces: Aiding User Critiques in Recommender Systems. *User Modeling*, 77–86.

Chen, L., and Pu, P. (2009). Interaction design guidelines on critiquing-based recommender systems. *User Modeling and User-Adapted Interaction*, *19*(3), 167.

Chen, L., and Pu, P. (2010). Experiments on the preference-based organization interface in recommender systems. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *17*(1), 1-33.

Chen, L., and Pu, P. (2012). Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction*, *22*(1–2), 125–150.

Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of consumer Research*, *30*(2), 170-183.

Cox, J., and Dale, B. G. (2001). Service quality and e-commerce: an exploratory analysis. *Managing Service Quality: An International Journal*.

Colucci, S., Noia, T. D., Sciascio, E. D., Donini, F. M., Ragone, A., and Rizzi, R. (2006). A semantic-based fully visual application for matchmaking and query refinement in B2C e-marketplaces. *Proceedings of the 8th international conference on Electronic commerce The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet, ICEC*, 174–184.

Crites, J. O. (1969). *Vocational Psychology The Study of Vocational Behavior and Dev*. McGraw-Hill.

Dalal, D. K., Diab, D. L., Zhu, X., and Hwang, T. (2015). Understanding the construct of maximizing tendency: A theoretical and empirical evaluation. *Journal of Behavioral Decision Making*, *28*(5), 437-450.

Dolfen, P., Einav, L., Klenow, P. J., Klopack, B., Levin, J. D., Levin, L., and Best, W. (2019). *Assessing the gains from e-commerce* (No. w25610). National Bureau of Economic Research.

Doyle, J. (1999). Rational decision making. *MIT encyclopedia of the cognitive sciences*, 701-703.

Ekstrand, M. D., Harper, F. M., Willemsen, M. C., and Konstan, J. A. (2014). User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender systems*, 161-168.

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2006). An Integrated Environment for the Development of Knowledge-Based Recommender Applications. *International Journal of Electronic Commerce*, *11*(2), 11–34.

Ferrari, J. R., and Dovidio, J. F. (2000). Examining behavioral processes in indecision: Decisional procrastination and decision-making style. *Journal of Research in Personality*, *34*(1), 127-137.

Frisch, D., and Clemen, R. T. (1994). Beyond expected utility: rethinking behavioral decision research. *Psychological bulletin*, *116*(1), 46.

Frost, R. O., and Shows, D. L. (1993). The nature and measurement of compulsive indecisiveness. *Behaviour Research and Therapy*, *31*(7), 683-IN2.

Grenci, R. T., and Todd, P. A. (2002). Solutionsdriven marketing. *Communications of the ACM,45*(2), 64-71.

Gudigantala, N., Song, J., and Jones, D. R. (2008). How well do e-commerce websites support compensatory and non-compensatory decision strategies? An exploratory study. *International Journal of E-Business Research (IJEBR)*, *4*(4), 43-57.

Harley, A, (2020). Compensatory vs Noncompensatory: 2 Decision-Making Strategies. Nielsen Norman Group. URL: https://www.nngroup.com/articles/compensatory-noncompensatory-decisions/

Häubl, G., and Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science, 19*(1), 4-21.

He, H. A., Greenberg, S., and Huang, E. M. (2010). One size does not fit all: applying the transtheoretical model to energy feedback technology design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 927-936.

Hogarth, R. (1987). *Judgment and choice* (2nd ed.). New York: John Wiley and Sons.

Hutton, R. J., and Klein, G. (1999). Expert decision making. Systems Engineering: The Journal of The International Council on Systems Engineering, 2(1), 32-45.

Jansen, S. J., Coolen, H. C., and Goetgeluk, R. W. (2011). *The measurement and analysis of housing preference and choice*. Springer Science and Business Media, *1*(1).

Jiang, Y., Shang, J., and Liu, Y. (2010). Maximizing customer satisfaction through an online recommendation system: A novel associative classification model. *Decision Support Systems*, *48*(3), 470-479.

Jin, Y., Cai, W., Chen, L., Htun, N. N., and Verbert, K. (2019). MusicBot: Evaluating Critiquing-Based Music Recommenders with Conversational Interaction. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 951–960.

Jugovac, M., and Jannach, D. (2017). Interacting with Recommenders—Overview and Research Directions. *ACM Transactions on Interactive Intelligent Systems*, *7*(3), 1–46.

Kamis, A., and Davern, M. J. (2004). Personalizing to product category knowledge: exploring the mediating effect of shopping tools on decision confidence. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the,* 10-pp.

Karimi, S., Papamichail, K. N., and Holland, C. P. (2015). The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour. *Decision Support Systems*, *77*, 137-147.

Willemsen, M. C., Graus, M. P., and Knijnenburg, B. P. (2016). Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction*, *26*(4), 347-389.

Knijnenburg, B. P., Reijmer, N. J. M., and Willemsen, M. C. (2011). Each to His Own: How Different Users Call for Different Interaction Methods in Recommender Systems. *Proceedings of the Fifth ACM Conference on Recommender Systems - RecSys '11*, 141–148.

Knijnenburg, B.P., Willemsen, M.C. (2009). Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system. In: Third ACM Conference on Recommender Systems, 381–384.

Knijnenburg, B.P.,Willemsen, M.C. (2010). The effect of preference elicitation methods on the user experience of a recommender system. In: 28th International Conference on Human Factors in Computing Systems, pp. 3457–3462.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, *22*(4–5), 441–504.

Knijnenburg, B. P., Willemsen, M. C., and Broeders, R. (2014). Smart Sustainability through System Satisfaction: Tailored Preference Elicitation for Energy-saving Recommenders. *AMCIS*, 1–15.

Knijnenburg, B. P., and Willemsen, M. C. (2015). Evaluating recommender systems with user experiments. In *Recommender Systems Handbook*, 309-352.

Konstan, J. A., and Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction*, *22*(1), 101-123.

Lai, K., and Cerpa, N. (2001). Support vs. confidence in association rule algorithms. In *Proceedings of the OPTIMA Conference, Curicó*, 1-14.

Langner, T., and Krengel, M. (2013). The mere categorization effect for complex products: The moderating role of expertise and affect. *Journal of Business Research*, *66*(7), 924-932.

Le Lec, F., Lumeau, M., and Tarroux, B. (2016). Choice or information overload? *Economics Working Paper Archive*, *7*.

Linden, G., Hanks, S., and Lesh, N. (1997). Interactive Assessment of User Preference Models: The Automated Travel Assistant. *User Modeling*, 67–78.

McCarthy, K., Reilly, J., McGinty, L., and Smyth, B. (2004). On the Dynamic Generation of Compound Critiques in Conversational Recommender Systems. *Lecture Notes in Computer Science*, *3137*, 176–184.

McGinty, L., and Smyth, B. (2003). On the Role of Diversity in Conversational Recommender Systems. *Case-Based Reasoning Research and Development*, 276–290.

McNeill, I. M., Dunlop, P. D., Skinner, T. C., and Morrison, D. L. (2015). Predicting delay in residents' decisions on defending v. evacuating through antecedents of decision avoidance. *International Journal of Wildland Fire*, *24*(2), 153-161.

Nenkov, G. Y., Morrin, M., Schwartz, B., Ward, A., and Hulland, J. (2008). A short form of the Maximization Scale: Factor structure, reliability and validity studies. *Judgment and Decision making*, *3*(5), 371-388.

Payne, J. W. (1982). Contingent Decision Behavior: A Review and Discussion of Issues. doi:10.21236/ada111655

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The Adaptive Decision Maker*, Cambridge, United Kingdom: Cambridge University Press, *105*.

Pereira, R. E. (2000). Optimizing human-computer interaction for the electronic commerce environment. *Journal of Electronic Commerce Research.*, *1*(1), 23–44.

Pu, P., Chen, L., and Kumar, P. (2008). Evaluating product search and recommender systems for E-commerce environments. *Electronic Commerce Research, 8*(1-2), 1-27.

Pu, P., and Faltings, B. (2004). Decision tradeoff using example-critiquing and constraint programming. *Constraints*, *9*(4), 289-310.

Rassin, E., Muris, P., Booster, E., and Kolsloot, I. (2008). Indecisiveness and informational tunnel vision. *Personality and Individual Differences*, *45*(1), 96-102.

Rassin, E., Muris, P., Franken, I., Smit, M., and Wong, M. (2007). Measuring general indecisiveness. *Journal of Psychopathology and Behavioral Assessment*, *29*(1), 60-67.

Randall, T., Terwiesch, C., and Ulrich, K. T. (2007). User Design of Customized Products. *Marketing Science*, *26*(2), 268–280.

Reilly, J., McCarthy, K., McGinty, L., and Smyth, B. (2004, August). Dynamic critiquing. In *European Conference on Case-Based Reasoning,* Springer, Berlin, Heidelberg, 763-777.

Reilly, J., McCarthy, K., McGinty, L., and Smyth, B. (2004, December). Incremental critiquing. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence,* Springer, London, 101-114.

McCarthy, K., McGinty, L., Smyth, B., and Reilly, J. (2005). A live-user evaluation of incremental dynamic critiquing. In *International Conference on Case-Based Reasoning,* Springer, Berlin, Heidelberg, 339-352.

Sechrest L. and Fay T.L. (1972) Problems of translation in cross-cultural research. *Journal of Cross-Cultural Psychology* 3, 41– 56.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129.

Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology, 41*(1), 1-20.

Schwartz, B. (2004). The paradox of choice: Why more is less. New York: Ecco.

Schwartz, B. (2016). On the meaning and measurement of maximization. *Judgment and Decision making*, *11*(2), 126.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., and Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, *83*(5), 1178.

Smyth, B., McGinty, L., Reilly, J., and McCarthy, K. (2004, September). Compound critiques for conversational recommender systems. In *IEEE/WIC/ACM International Conference on Web Intelligence,* 145-151.

TechTarget Contributors. (2020). association rules. Retrieved from https://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Xiao, and Benbasat. (2007). E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly, 31*(1), 137.

Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE transactions on knowledge and data engineering*, *12*(3), 372-390.

Zhou, L., Dai, L. and Zhang, D. (2007). Online shopping acceptance model-A critical survey of consumer factors in online shopping. *Journal of Electronic Commerce Research, 8*(1), 41.

Appendix 1: Questionnaires experiment

English	Dutch
Maximization tendency	
No matter what I do, I have the highest standards for myself.	Wat ik ook doe, ik stel altijd hoge eisen.
I never settle for second best.	Ik neem nooit genoegen met het op één na beste product.
No matter what it takes, I always try to choose the best option.	Ik doe er altijd alles aan om het beste product te vinden.
I don't like having to settle for "good enough."	Goed genoeg is voor mij niet voldoende.
I am a maximizer.	Ik probeer altijd het maximale resultaat te halen.
I will wait for the best option, no matter how long it takes.	Ik wacht tot ik het beste product vind, ongeacht hoe lang het duurt.
I never settle.	Ik sluit nooit een compromis wanneer ik een product wil aanschaffen.
Level of indecisiveness	
I find it easy to make decisions.	Ik vind het makkelijk om keuzes te maken.
After I have chosen something, I often believe I've made the wrong choice.	Na het maken van een keuze, twijfel ik of ik wel de juiste beslissing heb genomen.
Once I make a decision, I feel fairly confident that it is a good one.	Wanneer ik een keuze heb gemaakt, ben ik vrij zeker dat het de juiste keuze is.
I try to put off making decisions.	Ik stel het maken van keuzes vaak uit.
I usually make decisions quickly.	Meestal maak ik snel keuzes.
I like to be in a position to make decisions.	Ik vind het fijn om degene te zijn die verantwoordelijk is voor het maken van keuzes.
Perceived guidance[3]	
	De Keuzehulp maakt het makkelijker voor mij om te kiezen welke eigenschappen van de wasmachines ik belangrijk vind.
	De Keuzehulp motiveert me om betere afwegingen te maken tussen de wasmachines.
	De Keuzehulp is een eenvoudige manier om de best passende wasmachine voor mij te vinden.
	De Keuzehulp helpt me bij het begrijpen van de belangrijke eigenschappen van de wasmachines.
Choice satisfaction	
I like the product I've chosen.	Ik ben tevreden met de wasmachine die ik gekozen heb.
I think I chose the best product from the list.	Ik vind dat ik de beste wasmachine heb gekozen uit de lijst met opties.
The chosen product exactly fits my preference.	De door mij gekozen wasmachine past precies bij mijn voorkeuren.
I would rather choose another washing machine from the one I chose.	Ik had liever een andere wasmachine gekozen.
Perceived usefulness	
The system restricted my decision freedom.	Ik kon maar een beperkt aantal voorkeuren aangeven in de Keuzehulp.
I would use the system more often if possible.	Ik wil deze Keuzehulp vaker gebruiken, ook bij andere producten.
I make better choices with the system.	Met behulp van de Keuzehulp kan ik betere keuzes maken.
The system was useless.	De Keuzehulp was nutteloos.
I would recommend the system to others.	Ik zou de Keuzehulp aan anderen aanraden.
Using the system would enable me to accomplish tasks more quickly.	De Keuzehulp zorgt ervoor dat ik sneller keuzes kan maken tussen verschillende wasmachines.
Using the system would make it easier to find a product.	De Keuzehulp zorgt ervoor dat ik op eenvoudige wijze een wasmachine kan kiezen.

Table 5 Items in English and translated in Dutch

Appendix 2: Decision tree Keuzehulp

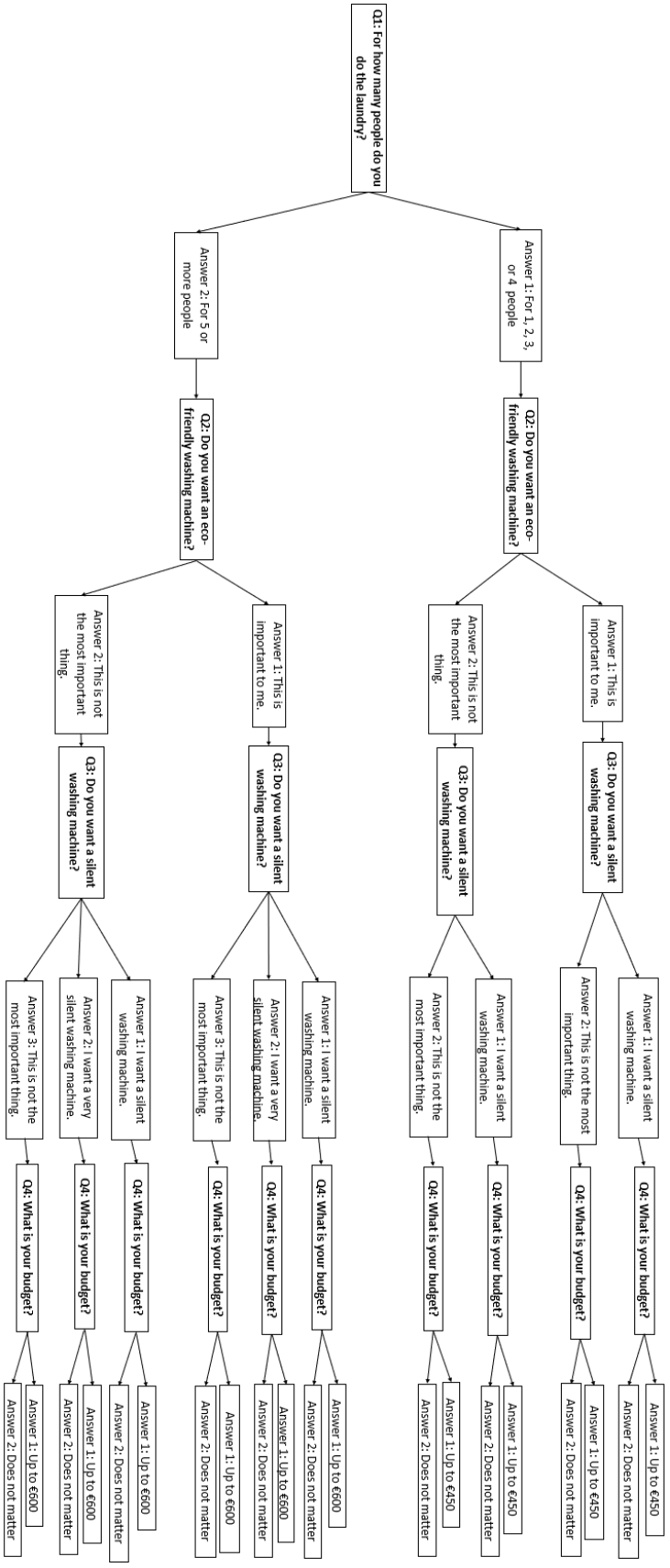

Fig. 13 Decision tree original Keuzehulp

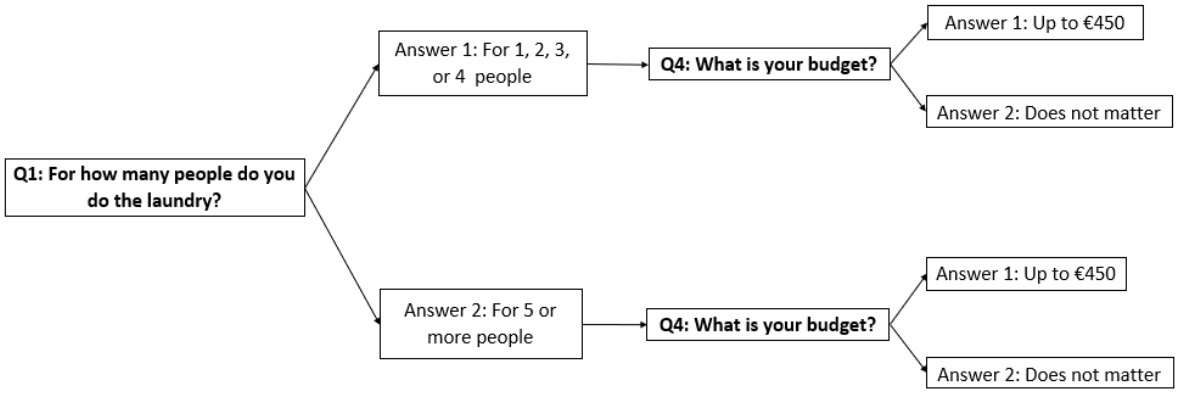

Fig. 14 Decision tree new Keuzehulp

Appendix 3: Design Keuzehulp

Original Keuzehulp

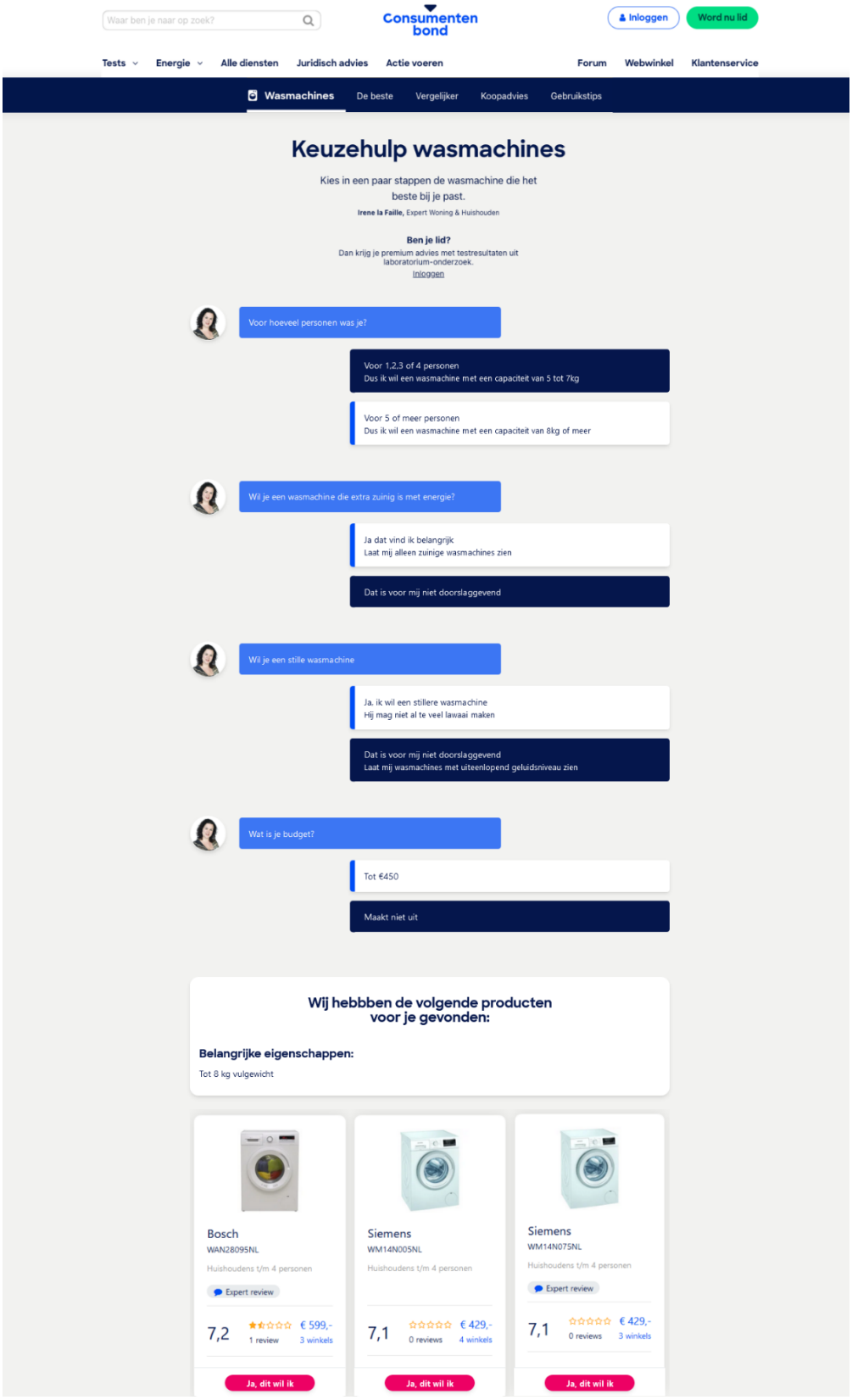

Fig. 15 Screenshot of example Original Keuzehulp.

New Keuzehulp

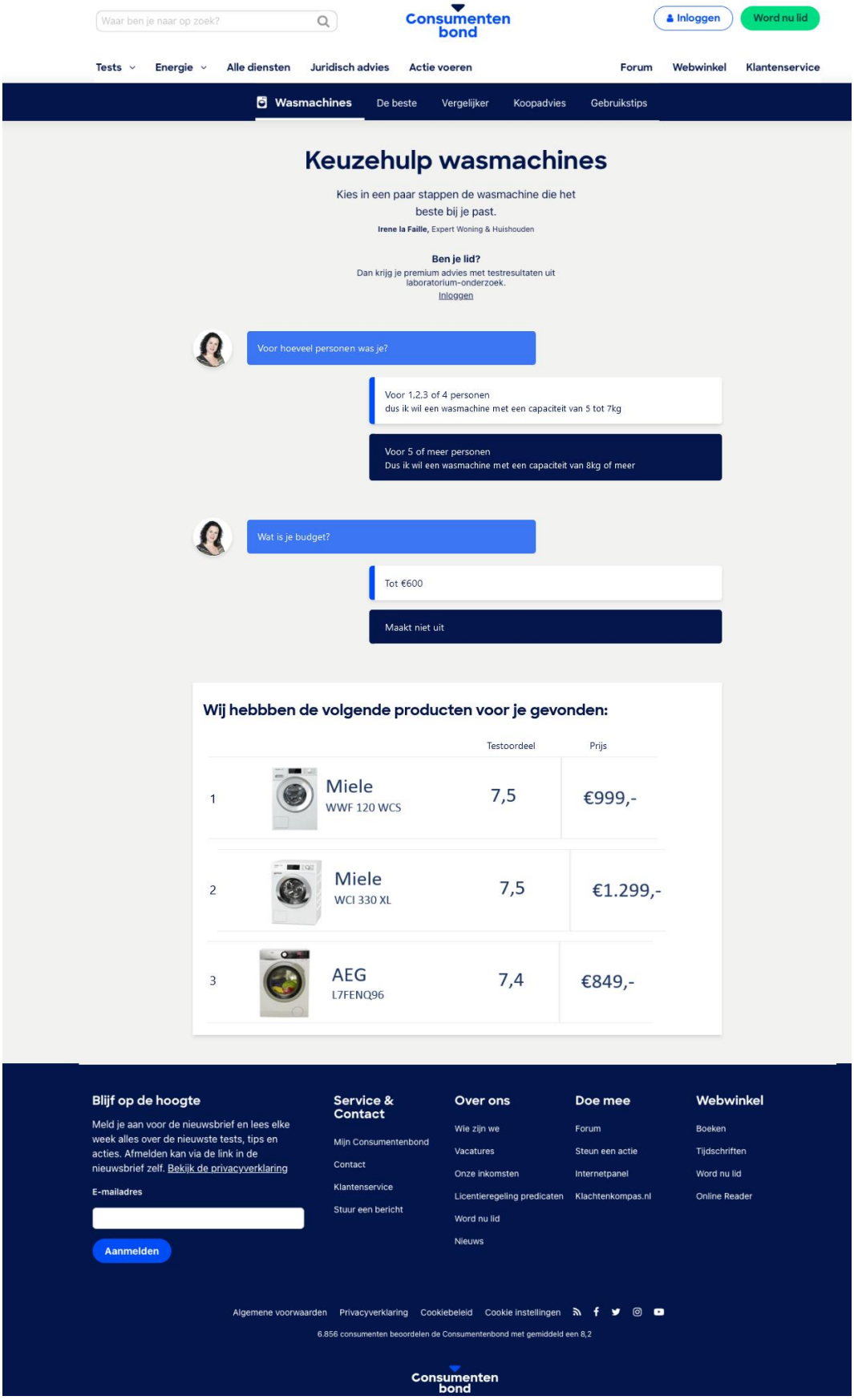

Fig. 16 Screenshot EBA (non-compensatory) part of the new Keuzehulp

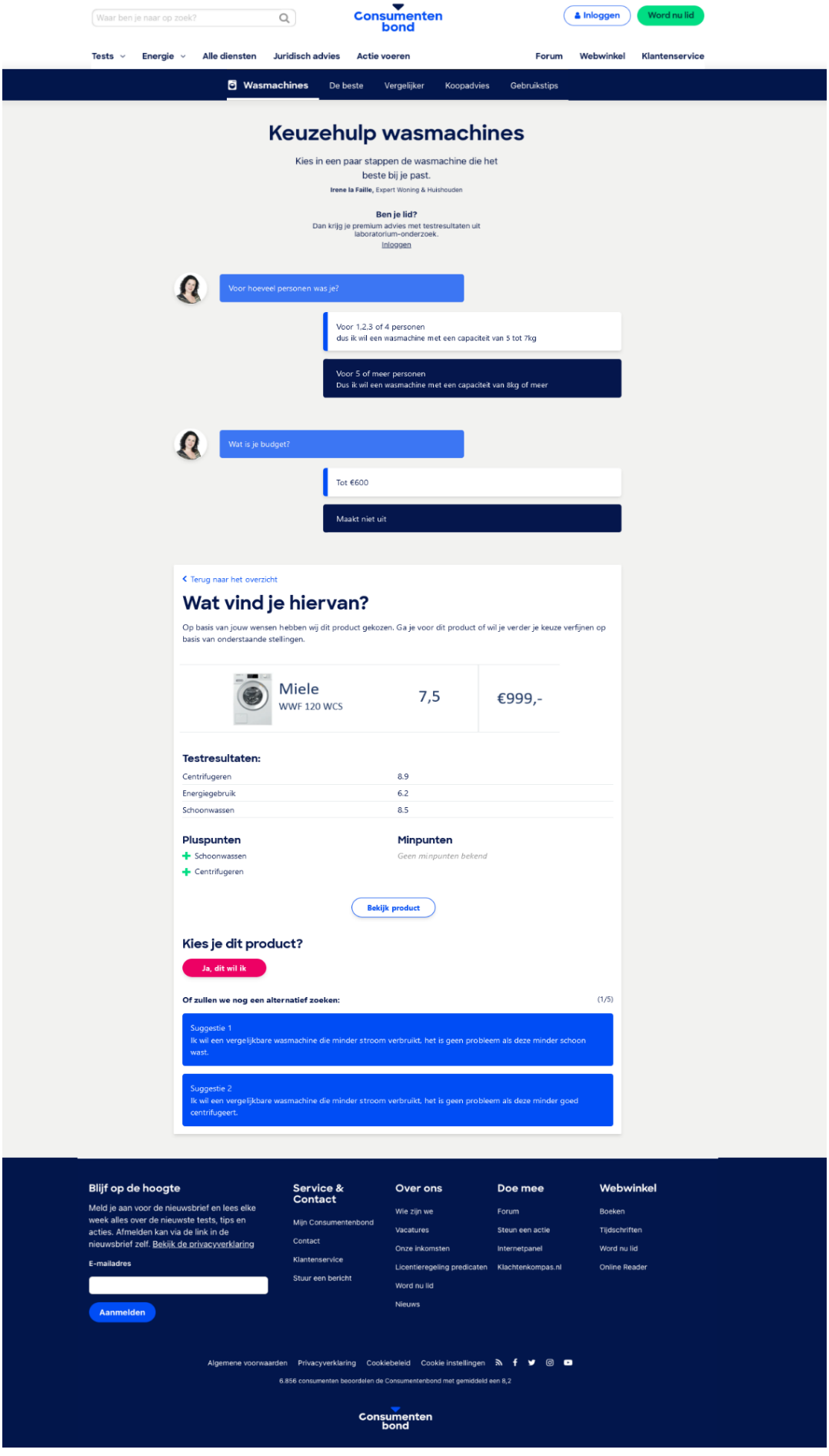

Fig. 17 Screenshot critiquing (compensatory) part of the new Keuzehulp

Appendix 4: Preference-based algorithm

Algorithm

Below, a more detailed explanation will be given regarding the implementation of the critiquing part of the newly developed Keuzehulp. The generation of critiques was adopted from a heuristic method that was validated by Chen and Pu (2007c).

Recommendation Algorithm

First, the interface asks the user a set of questions related to specific product specifications. The answers to these questions were translated into filter settings. After building the users' initial preference profile, the system generated a consideration set of three washing machines. These top three were based on their overall test scores, which were provided by the Consumentenbond. These average test scores were found by calculating the weighted addition of a products' test scores for its relevant attributes. The user could either choose to pick one of these three products or to critique one of the products.

Critiquing algorithm

Participants were given two system-suggested critiques for each product (e.g. "Ik wil een vergelijkbare wasmachine die beter spoelt, het is geen probleem als deze minder goed centrifugeert."). The personalized and diverse critiques were based on eight needs-based attributes (**Table 6**). These needs-based attributes were identified by experts from the Consumentenbond. In addition, the experts at the Consumentenbond assigned test scores and importance weights (**Table *6***) to each of the eight attributes for all products.

Table 6. Attributes and their importance weights

Product attributes	**Importance weights**
Cleanness ("Schoonwassen")	45%
Rinsing ("Spoelen")	10%
Centrifugation ("Centrifugeren")	10%
Eco-friendly ("Energieverbruik")	12%
Water usage ("Waterverbruik")	3%
Speed ("Snelheid")	5%
Ease of use ("Gebruiksgemak")	10%
Noise ("Geluid")	5%

The critiquing algorithm involved four steps. First, each product, except for the product that was being critiqued, was converted to a trade-off vector, which comprises of a set of (*attribute, trade-off*) pairs. The trade-off property as determined by the suggested direction of the attribute. It indicated whether a certain attributes' test score was *improved* ("Centrifugation, higher") or *compromised* ("Rinsing, lower"), when compared with the product that was being critiqued. Products could have the same score for a specific attribute. Therefore, these attributes were denoted as *equal* (e.g. "Cleanness_equal"). An example trade-off vector could be: {(" Cleanness, equal"), ("Rinsing, lower"), ("Centrifugation, higher"), ("Eco-friendly, equal"), ("Water usage, lower"), (Speed, higher"), ("Ease of use, higher"), ("Noise, lower")}.

Next, an Apriori algorithm was used to discover the recurring and representative subsets of the attributes', trade-off pairs (Agrawal, R., Imielinski, T., Swami, A., 1993). The reason for applying this Apriori algorithm is due to its popularity and efficiency in mining associate rules among features (Pu and Chen, 2006).

Association rule mining

Association rules are used to find co-occurrences and correlations between datasets (TechTarget Contributors, 2020). This technique attempts to find groups of items that are commonly found together (Lai and Cerpa, 2001). Association rules are "if-then" statements, that show the probability of relationships between data items, within large datasets. At the basic level, association rule mining involves the use of machine learning models to try and analyse co-occurrences or patterns in a dataset. An association rule consists out of an antecedent and a consequent. The antecedents (if) is an item found within a dataset. The consequent (then) is an item found in combination with the antecedent. These association rules were created by searching the data for frequent if-then patterns, also called an itemset {antecedent, consequent}. Essentially, the association is about discovering a set of rules shared by a significant percentage of the data (Zaki, 2000). Often a large number of rules are produced when using association rule mining. The goal is to find rules that are useful to users. These rules are found by applying the criteria support, to identify the most important relationships. Support indicates the frequency that both the antecedent and the consequent both appear in the data. Mathematically, support is the fraction of the total number of products in which the itemset occurs. Both X and Y are in this case attributes. The minimum support that was used in this study is 0.33. This minimum support is chosen because we always wanted to recommend a minimum of 3 products to the users.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Products\ containing\ both\ X\ and\ Y}{total\ numbers\ of\ products}$$

The rules that were used in this study consisted out of two attributes. Using more attributes could have resulted into information overload (Le Lec, Lumeau, and Tarroux, 2016). Therefore each critique contained two attributes (*eg. "I would like a similar washing machine but that rinses the clothes better and it's no problem if it is less silent").*

After applying the Apriori algorithm, all products were placed into categories. Where each category represented a critique (e.g. "rinses better", "*less silent*"). Products could belong to more than one category because they could have had different subsets of trade-off properties shared by other groups of products. Therefore, a large number of critiques can be generated. To be able to select the most prominent critique, critiques were ranked according to their trade-off utilities. In other words, looking at their gains and losses relative to the top candidate, in terms of both critiques themselves and their contained products. By using this method, users' preferences were taken into account.

$$Trade\ off\ Utility(C) = (\sum_{i=1}^{|C|} w(attribute_i) * tradeoff_i) * (\frac{1}{|SR(C_i)|}\sum_{R \in SR(C_i)}^{|SR(C_i)|} U(r)) \quad (1)$$

The *C* denotes the critiques as a set of *attribute, trade-off* pairs and *SC(C)* is the set of products that satisfy *C*. The first part of the formula, $(\sum_{i=1}^{|C|} w(attribute_i) * tradeoff_i)$, computes the weighted sum of the trade-off properties. Where $w(attribute_i)$ is the weight of each $attribute_i$. During the first critiquing round, little is known about the users' preferences. Therefore a set of default weights were used (table..). These weights were chosen based on the expertise of the Consumentenbond. For $tradeoff_i$, a default was set as 0.25 if an attribute was compromised, or 0.75 if improved. These default settings were chosen because improved attributes are in nature more valuable than compromised attributes (Chen and Pu, 2007). The second part of the formula, $(\frac{1}{|SR(C_i)|}\sum_{R \in SR(C_i)}^{|SR(C_i)|} U(r))$, calculated the average utility of all products contained by *C*. The U denotes the utility of each product. User preferences are represented as a weighted addition of its values for its relevant attributes, according to the multi-attribute utility theory (MAUT) (Jansen, Coolen, and Goetgeluk, 2011). This theory follows the common and compensatory decision strategy, called the weighted additive rule. This resolves conflicting values explicitly by considering all trade-offs (Payne, Bettman, and Johnson, 1993). Where u_i is the value function for each attribute x_i, and w_i is the relative importance weight of x_i. The total utility of each product can be calculated as :

$$u(x_1, \dots, x_n) = \sum_{i=1}^{n} w_i u_i(x_i) \quad (2)$$

Next, the proposed critiques and their products were diversified to increase their suggestion power, since similar items have limited added value (McGinty and Smyth, 2003). The diversity

degree of C was calculated as the minimal local diversity of C with each critique Ci in the SC set. To define the local diversity of two critiques, two factors were important. The first factor calculates the diversity between critiques themselves (i.e. *C* and C_i). The second factor calculates the diversity between the products contained in each critique (i.e. *SC(C)* and $SC(C_i)$):

$$Diversity(C, SC) = \min_{C_i \in SC}\left(\left(1 - \frac{|C \cap C_i|}{|C|}\right) * \left(1 - \frac{|SC(C) \cap SC(C_i)|}{|SC(C)|}\right)\right) \quad (3)$$

The trade-off utility of each critique is then multiplied by a diversity degree:

$$F(C) = TradeoffUtility(C) * Diversity(C, SC) \quad (4)$$

The degree of diversity denoted the minimal local diversity of *C* with each critique C_i in the SC set. Where *SC* denotes the set of critiques that are selected so far. The top critique was hence the critique with the highest trade-off utility, and the second chosen critique that was shown to the users was the critique with the highest value of F(C) in the remaining non-selected critiques. The selection process ended when the desired critiques were determined.

After the two top critiques were generated, users were able to select one of the two proposed critiques. When one of the critiques was chosen a new consideration set was suggested to the user. The top three products were chosen based on their average test score. These average test scores were calculated based on MAUT as discussed earlier. The relative importance weights used to calculate the average test scores were refined after every critiquing cycle. As discussed before, during the first cycle, the importance weights were based on the weights assigned by the Consumentenbond (Table 1). During the subsequent critiquing cycles, the importance weights were updated, based on users' critiquing history. Improved attributes that were discussed in the selected critique were increased by β. When an attribute was compromised, the importance weight decreased by β, whereby **β=0.02**. For example, "*I would like a similar washing machine but that rinses the clothes better and it's no problem if it is less silent*". The importance weight for the attribute "Rinsing" would increase with 0.02 and the attribute "Noise" would decrease with 0.02. The value 0.02 was chosen with the aim of gradually optimizing the importance of the attributes for each user over time, without drastically changing the nature of the recommendations with each critique[2]. In addition, several attributes had a relatively low importance weight to begin with (e.g. "Water usage" had an importance weight of 3%). If a higher weight was chosen, users would never have the opportunity to critique these attributes. Due to the public nature of this study, users were able to critique products up to five times. The critiques became more personalized when more information about the users was collected.

[2]Not Due to time constraints, the optimal values for optimizing the weights were not empirically validated.

Appendix 5: Statistical output

Maximization tendency and level of indecisiveness

Table 7 Mean choice satisfaction and perceived guidance for satisficers (Low) and maximizers (High). SE indicated standard error of the mean

	Choice satisfaction Mean (SE)		Perceived guidance Mean (SE)	
Maximization	Original Keuzehulp	New Keuzehulp	Original Keuzehulp	New Keuzehulp
Low	-0.04 (0.08)	-0.04 (0.10)	0.00 (0.08)	-0.15 (0.10)
High	-0.14 (0.09)	0.21 (0.09)	-0.01 (0.09)	0.13 (0.10)

Table 8 Mean choice satisfaction and perceived guidance for participants who are not indecisive (Low) and participants who are indecisive (High). SE indicated standard error of the mean

	Choice satisfaction Mean (SE)		Perceived guidance Mean (SE)	
Level of indecisiveness	Original Keuzehulp	New Keuzehulp	Original Keuzehulp	New Keuzehulp
Low	-0.30 (0.08)	-0.06 (0.09)	-0.01 (0.08)	-0.13 (0.10)
High	0.11 (0.08)	0.24 (0.10)	0.00 (0.08)	0.12 (0.10)

Goodness of fit SEM model

```
lavaan 0.6-7 ended normally after 28 iterations

  Estimator                                         DWLS
  Optimization method                           NLMINB
  Number of free parameters                         58

  Number of observations                           434

Model Test User Model:
                                              Standard      Robust
  Test Statistic                                37.257      88.113
  Degrees of freedom                                40          40
  P-value (Chi-square)                           0.594       0.000
  Scaling correction factor                                  0.473
  Shift parameter                                            9.326
       simple second-order correction

Model Test Baseline Model:

  Test statistic                             17740.919    8618.915
  Degrees of freedom                                21          21
  P-value                                        0.000       0.000
  Scaling correction factor                                  2.061

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    1.000       0.994
  Tucker-Lewis Index (TLI)                       1.000       0.997

  Robust Comparative Fit Index (CFI)                            NA
  Robust Tucker-Lewis Index (TLI)                               NA

Root Mean Square Error of Approximation:

  RMSEA                                          0.000       0.053
  90 Percent confidence interval - lower         0.000       0.038
  90 Percent confidence interval - upper         0.030       0.068
  P-value RMSEA <= 0.05                          1.000       0.362

  Robust RMSEA                                                  NA
  90 Percent confidence interval - lower                        NA
  90 Percent confidence interval - upper                        NA

Standardized Root Mean Square Residual:

  SRMR                                           0.026       0.026
```

Covariances

```
                     Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.al
  Maximization ~~
    Indecisiveness     -0.072    0.049   -1.469    0.142   -0.072   -0.072
    Perceivedgudnc      0.057    0.049    1.152    0.249    0.057    0.057
    Choicesatsfctn      0.044    0.049    0.912    0.362    0.044    0.044
  Indecisiveness ~~
    Perceivedgudnc      0.092    0.051    1.817    0.069    0.092    0.092
    Choicesatsfctn      0.189    0.050    3.746    0.000    0.189    0.189
  Perceivedguidance ~~
    Choicesatsfctn      0.634    0.028   22.292    0.000    0.634    0.634
```

Modification indices

```
                lhs op   rhs      mi    epc sepc.lv sepc.all sepc.nox
       Maximization =~ IND_6 188.476  0.321   0.321    0.321    0.321
       Maximization =~ IND_3 158.277  0.312   0.312    0.312    0.312
 Choicesatisfaction =~ USE_1 122.467 -0.496  -0.496   -0.496   -0.496
              IND_2 ~~ IND_3  74.567 -0.330  -0.330   -0.600   -0.600
                M_5 ~~ IND_6  71.964  0.301   0.301    0.539    0.539
                M_5 ~~ IND_3  68.461  0.312   0.312    0.627    0.627
              USE_6 ~~ USE_7  66.889  0.139   0.139    0.580    0.580
     Indecisiveness =~   M_5  66.119  0.227   0.227    0.227    0.227
              IND_1 ~~ IND_5  63.186  0.330   0.330    0.798    0.798
       Maximization =~ IND_5  61.680 -0.207  -0.207   -0.207   -0.207
              USE_2 ~~ USE_3  52.606  0.108   0.108    0.646    0.646
              IND_3 ~~ IND_5  50.135 -0.348  -0.348   -0.667   -0.667
       Maximization =~ IND_1  46.379 -0.189  -0.189   -0.189   -0.189
       Maximization =~ IND_2  40.066  0.162   0.162    0.162    0.162
       Choice_sat_2 ~~ USE_1  35.390 -0.218  -0.218   -0.708   -0.708
  Perceivedguidance =~ IND_2  35.342  0.111   0.111    0.111    0.111
       Choice_sat_1 ~~ USE_1  33.860 -0.216  -0.216   -0.744   -0.744
Perceivedusefulness =~ IND_2  33.860 -0.105  -0.105   -0.105   -0.105
                M_3 ~~ IND_3  32.741  0.241   0.241    0.466    0.466
                M_3 ~~ IND_6  31.370  0.246   0.246    0.423    0.423
```

SEM model output

Regressions:

```
                     Estimate  Std.Err  z-value  P(>|z|)
  Choice.satisfaction ~
    Keuzehulp           0.357    0.110    3.248    0.001
    Perceived.gdnc      0.818    0.059   13.894    0.000
    ind_C               0.227    0.058    3.884    0.000
    KHC_IND            -0.168    0.115   -1.466    0.143
  Perceived.guidance ~
    Keuzehulp          -0.025    0.103   -0.241    0.810
    max_C               0.077    0.057    1.356    0.175
    KHC_MAX             0.230    0.114    2.023    0.043
    ind_C               0.129    0.054    2.382    0.017
    KHC_IND             0.197    0.109    1.811    0.070
```

Moderated mediation

Maximization

Regressions:

```
                     Estimate  Std.Err  z-value  P(>|z|)
```

```
  Choice.satisfaction ~
    Keuzehlp_C              0.357    0.110    3.248    0.001
    Prcvd.gdnc (b)          0.818    0.059   13.870    0.000
    ind_C                   0.227    0.058    3.885    0.000
    max_C                   0.032    0.050    0.648    0.517
    KHC_IND                -0.168    0.115   -1.465    0.143
    KHC_MAX    (c)          0.004    0.099    0.041    0.968
  Perceived.guidance ~
    Keuzehlp_C             -0.025    0.103   -0.241    0.810
    ind_C                   0.129    0.054    2.382    0.017
    max_C                   0.066    0.055    1.198    0.231
    KHC_MAX    (a)          0.229    0.111    2.068    0.039
    KHC_IND                 0.197    0.109    1.811    0.070

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ab                0.187    0.091    2.056    0.040
```

Level of indecisiveness

```
Regressions:
                     Estimate  Std.Err  z-value  P(>|z|)
  Choice.satisfaction ~
    Keuzehlp_C              0.357    0.110    3.248    0.001
    Prcvd.gdnc (b)          0.818    0.059   13.870    0.000
    ind_C                   0.227    0.058    3.885    0.000
    max_C                   0.032    0.050    0.648    0.517
    KHC_IND    (c)         -0.168    0.115   -1.465    0.143
    KHC_MAX                 0.004    0.099    0.041    0.968
  Perceived.guidance ~
    Keuzehlp_C             -0.025    0.103   -0.241    0.810
    ind_C                   0.129    0.054    2.382    0.017
    max_C                   0.066    0.055    1.198    0.231
    KHC_MAX                 0.229    0.111    2.068    0.039
    KHC_IND    (a)          0.197    0.109    1.811    0.070

Defined Parameters:
                   Estimate  Std.Err  z-value  P(>|z|)
    ab                0.161    0.089    1.808    0.071
```