Decision-Making Support in Complex Multi-Actor and Multi-Source Scenarios

Faculty of Engineering

Department of Computer Science and Applied Cognitive Science

of the University of Duisburg-Essen

for the attainment of the academic degree of

Doctor of Engineering (Dr.-Ing.)

approved cumulative dissertation of

Jesús Omar Álvarez Márquez

from

Granada, Spain

First Supervisor: Prof. Dr.-Ing. Jürgen Ziegler

Second Supervisor: Prof. Dr. rer. nat. Mohamed Amine Chatti

Date of the oral exam: 14.12.2022

ABSTRACT

The term "complex decision environment" has been used in literature to describe situations that involve making a decision with one or more of the following aspects: multiple stakeholders, multiple viable alternatives, contradictory or incomplete information, multiple sources, and dynamic settings. This thesis aims at achieving a better understanding of the processes that take place in or under complex decision environments, with a focus on cases where multiple actors or multiple sources of information are present. To that end, the first goal of this research work is to define specific but sufficiently common scenarios in which multi-factor decision-making takes place. Examining these scenarios can lead to a greater comprehension of the particular requirements, decision processes and challenges faced by decision-makers when caught in a multi-factor decision environment. After defining these scenarios and their characteristics, a second research goal is to design and evaluate suitable methods to support the decision-making process that takes place in them.

The thesis presents investigations and developments in two different application contexts. The first one concentrates on complex multi-actor decision environments, illustrated by a group of non-collocated people who must jointly choose a hotel to stay together. The second scenario deals with complex multi-source decision environments, and more specifically with consumers facing a purchase decision in a physical store setting where online information is also available. Methods have been designed to assist decision-makers by taking into account the unique characteristics of each scenario, for which the use of recommender systems and technological innovations (especially in relation to augmented reality) has proven to be advantageous. The evaluation of the developed methods provides insight into the cognitive processes and constraints that emerge in multifactorial settings, and establishes design guidelines for future research.

Keywords: complex decision environment, multiple actors, multiple sources, group recommender system, negotiation of preferences, in-store shopping support, augmented reality

ZUSAMMENFASSUNG

Der Begriff “komplexes Entscheidungsumfeld” wird in der Literatur verwendet, um Situationen zu beschreiben, in denen eine Entscheidung mit einem oder mehreren der folgenden Aspekte getroffen werden muss: mehrere Akteure, mehrere realisierbare Alternativen, widersprüchliche oder unvollständige Informationen, mehrere Quellen und dynamische Rahmenbedingungen. Ziel dieser Arbeit ist es, ein besseres Verständnis der Prozesse zu erlangen, die in oder unter komplexen Entscheidungsumgebungen ablaufen, wobei der Schwerpunkt auf Fällen liegt, in denen mehrere Akteure oder mehrere Informationsquellen vorhanden sind. Zu diesem Zweck besteht das erste Ziel dieser Forschungsarbeit darin, spezifische, aber hinreichend häufige Szenarien zu definieren, in denen eine multifaktorielle Entscheidungsfindung stattfindet. Die Untersuchung dieser Szenarien kann zu einem besseren Verständnis der besonderen Anforderungen, Entscheidungsprozesse und Herausforderungen führen, mit denen sich Entscheidungsträger in einem multifaktoriellen Entscheidungsumfeld konfrontiert sehen. Nach der Definition dieser Szenarien und ihrer Charakteristika besteht ein zweites Forschungsziel darin, geeignete Methoden zur Unterstützung des in ihnen ablaufenden Entscheidungsprozesses zu entwickeln und zu evaluieren.

In dieser Arbeit werden Untersuchungen und Entwicklungen in zwei verschiedenen Anwendungskontexten vorgestellt. Das erste Szenario konzentriert sich auf komplexe Multi-Akteurs-Entscheidungsumgebungen, illustriert durch eine Gruppe von nicht kollokierten Personen, die sich gemeinsam für ein Hotel entscheiden müssen, in dem sie übernachten wollen. Das zweite Szenario befasst sich mit komplexen Entscheidungsumgebungen mit mehreren Quellen, genauer gesagt mit Verbrauchern, die vor einer Kaufentscheidung in einem physischen Geschäft stehen, in dem auch Online-Informationen verfügbar sind. Es wurden Methoden entwickelt, um Entscheidungsträger zu unterstützen, indem die einzigartigen Merkmale jedes Szenarios berücksichtigt werden, wofür sich der Einsatz von Empfehlungssystemen und technologischen Innovationen (insbesondere in Bezug auf Augmented Reality) als vorteilhaft erwiesen hat. Die Evaluierung der entwickelten Methoden bietet Einblicke in die kognitiven Prozesse und Einschränkungen, die in multifaktoriellen Umgebungen auftreten, und legt Design-Richtlinien für zukünftige Forschung fest.

Keywords: komplexe Entscheidungsumgebung, mehrere Akteure, mehrere Quellen, Gruppenempfehlungssystem, Aushandlung von Präferenzen, Einkaufsunterstützung in Geschäften, erweiterte Realität

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Jürgen Ziegler, first of all for giving me the opportunity to be part of the Interactive Systems Research Group and, second, for his invaluable support during these years of work. His knowledge, experience and guidance have been fundamental not only for the content of this thesis, but also for my personal and professional development.

I would like to extend my deepest gratitude to my friends and colleagues in Germany. This journey would not have been possible without their warm welcome and hospitality during all these years away from home.

Last but not least, I would like to thank the two most important women in my life, my mother Gloria and my wife Berta, whose unconditional support and love motivate me to always try to be my best self.

Table of Contents

List of Papers Included in this Dissertation

Paper 1 Álvarez Márquez, J. O. & Ziegler, J. (2015). *Preference Elicitation and Negotiation in a Group Recommender System.* In *Human-Computer Interaction – INTERACT 2015*, Lecture Notes in Computer Science, pp. 20–37. Springer International Publishing. https://doi.org/10.1007/978-3-319-22668-2_2

Paper 2 Álvarez Márquez, J. O. & Ziegler, J. (2016). *Hootle+: A Group Recommender System Supporting Preference Negotiation*. In *Collaboration and Technology*, Lecture Notes in Computer Science, pp. 151–166. Springer International Publishing. https://doi.org/10.1007/978-3-319-44799-5_12

Paper 3 Álvarez Márquez, J. O. & Ziegler, J. (2018a). *Negotiation and Reconciliation of Preferences in a Group Recommender System*. Journal of Information Processing, 26:186–200. https://doi.org/10.2197/ipsjjip.26.186

Paper 4 Álvarez Márquez, J. O. & Ziegler, J. (2018b). *Augmented Reality Based Recommending in the Physical World*. In *Mensch und Computer 2018-Workshopband*. Gesellschaft für Informatik eV. https://doi.org/10.18420/muc2018-ws07-0461

Paper 5 Álvarez Márquez, J. O. & Ziegler, J. (2019). *Augmented-Reality-Enhanced Product Comparison in Physical Retailing*. In *Proceedings of Mensch und Computer 2019*, MuC'19, pp. 55–65. Association for Computing Machinery. https://doi.org/10.1145/3340764.3340800

Paper 6 Álvarez Márquez, J. O. & Ziegler, J. (2020). *In-Store Augmented Reality-Enabled Product Comparison and Recommendation*. In *Fourteenth ACM Conference on Recommender Systems*, RecSys '20, pp. 180–189. Association for Computing Machinery. https://doi.org/10.1145/3383313.3412266

Paper 7 Álvarez Márquez, J. O. & Ziegler, J. (2021). *Acceptance of an AR-Based In-Store Shopping Advisor - the Impact of Psychological User Characteristics*. In *Human-Computer Interaction – INTERACT 2021*, Lecture Notes in Computer Science, pp. 457–479. Springer International Publishing. https://doi.org/10.1007/978-3-030-85623-6_28

Paper 8 Álvarez Márquez, J. O. & Ziegler, J. (2023). *Creating Omni-Channel In-Store Shopping Experiences through Augmented-Reality-Based Product Recommending and Comparison*. International Journal of Human–Computer Interaction, 1-26. https://doi.org/10.1080/10447318.2022.2163650

List of Abbreviations

AR Augmented Reality

GDM Group Decision Making

GRS Group Recommender System

HMD Head Mounted Display

IoT Internet of Things

RS Recommender System

List of Tables

LIST OF FIGURES

1
INTRODUCTION

It is often said that *"good information leads to good decision-making"*. The statement seems to be close to reality, as early research has shown that reaching a satisfactory decision highly depends on the availability and quality of pertinent information, while irrelevant one may instead hinder the decision-making process (O'Reilly III, 1982). Hence, an ideal decision environment is that which allows easy access to all relevant information about all possible alternatives, so that a decision-maker is able to identify the best possible choice. However, perfect information is rarely the case, and usually a decision must be reached by using unreliable or incomplete data (Simon, 1997; Nielsen, 2011). Any obstacle in the way of reaching a decision is also further magnified when multiple decision-makers or information sources are considered. In this sense, research on decision-making support has been most notable in the form of decision support systems (Arnott & Pervan, 2015) and recommender systems (Ricci et al., 2011). The former approach offers methods for multi-criteria decision-making and the creation of information reports by collecting and analysing large amounts of data, but these solutions are mostly developed for their use in business scenarios with organizational or managerial purposes. Recommender systems (RS), on the other hand, place more emphasis on user preferences and their utilization in the generation and presentation of a reduced number of relevant alternatives, and may be better suited in a broader range of multi-actor and multi-source decision environments. Nonetheless, new technology advances have resulted in the appearance of more complex scenarios where decision making can take place, for which existing research may not be sufficient.

Concerning multi-actor decision scenarios, people are more connected today than ever, but traditional face-to-face communication has been displaced in favour of less direct conversational methods (Meier et al., 2021; Shufford et al., 2021). Although voice and video calls are widely used, it is still more common to communicate via e-mail or chat applications in most everyday situations (Romiszowski & Mason, 2013), more so when the circumstances require for the persons involved to research and think about a matter by themselves, or when scheduling a real-time meeting is not possible. Using e-mail services or chat applications often implies asynchronous communication, which precisely permits such kind of individual thinking and time organization (Berry, 2006); per contra, it also means that most relevant aspects of face-to-face group dynamics are lost, and the intent of a message can often be misinterpreted and errors may be more frequent (Damian & Zowghi, 2002; Marlow et al., 2017). These limitations may increase the complexity of multi-actor decision-making scenarios where each individual has different preferences and access possibilities to information, such as when collectively deciding on a film, a hotel, or a sightseeing route. In these cases, reaching consensus on which alternative to choose may become a long and convoluted process, at the end of which the likelihood of dissatisfaction of some involved actors may escalate due to communication limitations. These complex multi-actor decision environments may benefit from group recommender systems (GRS), as they provide recommendations that match the preferences of multiple users (Felfernig et al., 2018). However, current GRS do not offer enough flexibility and user agency during the recommending process to sufficiently support group decision-making, which would require the inclusion of appropriate preference negotiation and consensus building methods.

Similarly, new technology advances allow for information to be accessible at any time and place (Church et al., 2007; Hilbert & López, 2011). The growing ubiquity of information, which in principle should ease the process of making a choice by providing useful data about different alternatives (Streufert, 1973; Citroen, 2011; Haas et al., 2015), can also become an issue when decision-makers are left alone to collect and filter information in a setting where many relevant heterogeneous sources have to be considered. In addition to the demanding task of searching for, interpreting and selecting information, it is also important to observe that some of these sources may not even belong to the same medium, at which point integrating all the relevant knowledge becomes a complicated process (Wolny & Charoensuksai, 2014). As an example, this is the case for physical shopping scenarios, where product data is not only accessible through sheets, posters, or sales personnel, but it also may have an experiential origin through direct product inspection, and consumers may as well resort to external online resources (Broeckelmann & Groeppel-Klein, 2008). Such amalgam of information origins may result in a very complex decision environment, and risk exists for issues like information overload or mental fatigue to appear, which amplify the possibilities of poor decision-making (Chen et al., 2009; Mix & Katzberg, 2015; Lu & Gursoy, 2015). RS can also be beneficial in multi-source decision situations, as they enable a greater abstraction of the sources of information by displaying data retrieved from them in a more cohesive manner in the shape of recommended items. Still, research is scarce when it comes to the joint presentation of recommendations that make use of physical and digital sources, not only in terms of data visualization, but also concerning the implications that such hybrid environment could have in the way users navigate and evaluate information.

Complex multi-factor decision environments such as these are not unusual today, and will become more common as information technologies advance. Conveniently, the same technology that has originated these new complex scenarios provides the means through which their limitations can be alleviated, and also opens new possibilities for supporting them by applying RS in novel ways. For instance, modern communication technologies allow the design of platforms to support group decision-making supported by GRS, where group members can express and negotiate their preferences on a matter by using more flexible methods than a simple text chat (Alvarado Rodriguez et al., 2022). Moreover, technologies like augmented reality (AR) are revolutionizing the way by which information is accessed, displayed and interacted with, and open new opportunities for creating mixed environments where recommendations using information from multiple sources (physical or digital) can be presented together (Cruz et al., 2019; Gutiérrez et al., 2019; Ludwig et al., 2020).

The search for methods to reduce uncertainty and support decision-makers has been widely investigated in social and formal science fields (Eom, 1997; Arnott & Pervan, 2015), and literature is extensive for most of the different areas discussed here. Despite this, current research has mostly overlooked these type of complex decision environments where providing support is not trivial. However, as technology enables the increasingly frequent emergence of multi-actor and multi-source scenarios, a more in-depth study of their specific characteristics and challenges is required, along with the investigation of new methods for providing decision support in these situations. As already mentioned, a promising research direction lies in the use of RS, since they are capable to integrate information and reduce the effort that decision-makers must do to analyse it, all while taking into account the preferences of one or multiple actors. However, investigating appropriate methods for the application of RS in these particular scenarios is still required.

This dissertation is structured as follows: chapter 2 presents the background and related work; chapter 3 describes the research goals of the thesis and outlines the specific application settings on which it focuses; chapter 4 summarizes the contributions of each paper included in this dissertation; and chapter 5 discusses the conclusions of this research. Lastly, a full reproduction of each contributing paper can be found in the appendix.

2 Theoretical Background

2.1 The Decision-Making Process

Decision-making has been identified as one of the fundamental cognitive processes of human behaviour (Wang et al., 2006). It is often defined as *identifying and choosing alternatives based on the values and preferences of the decision-maker* (Harris, 1998). This implies that multiple alternative choices must exist for a decision to be made, from which the one that is likely to provide the most benefit and better fits certain preferences is to be chosen.

Despite extensive research, decision-making is an area still full of competing models and theories, in part due to the multidisciplinary nature of the subject, which can be approached from fields as varied as psychology, economics, sociology, computer science, mathematics or cognitive science (Harrison, 1999; Wald, 1950; Edwards & Fasolo, 2001; Wang et al., 2006; Far & Wahono, 2003). However, historically two basic models are emphasized, the *rational model* and the *bounded rationality model* (Lunenburg, 2010a).

In the rational model, decision-making is seen as a logical sequence of activities performed under certainty, which may be repeated an unlimited number of times in an iterative manner, until a satisfactory decision is made. The whole process can be broken into six stages (Schoenfeld, 2010):

1. Identifying the problem: the goals of the decision-making process are established by finding the problems that have to be solved. Identifying the problem is crucial for deciding a course of action and evaluating the decision outcome.
2. Generating alternatives: the number of alternatives should be as large and diverse as possible, but the search is limited by the importance of the decision, and the cost and value of additional information (Zopounidis & Pardalos, 2010).
3. Evaluating alternatives: for each alternative, it must be analysed its feasibility, to which extent it addresses the problem, and the consequences of choosing it.
4. Choosing an alternative: alternatives are compared and the most satisfactory one is chosen. The comparison should be made in terms of the degree to which their outcomes and consequences achieve the desired objectives.

5. Implementing the decision: the chosen alternative is used to address the original problem.
6. Evaluating decision effectiveness: the outcome of implementing the chosen alternative is evaluated to check the extent to which it produces the desired effects. Depending on the effectiveness of the decision, it may be required to make a new one, usually after revisiting the analysis of the problem, generating new alternatives, or selecting a different one.

Some authors argue that having perfect information transforms the decision-making process into a matter of optimization, where the choice of the most satisfactory solution becomes "trivial". In real world scenarios, neither decision-makers are completely rational nor full information is usually available. Because our environment is not a deterministic place, uncertainty becomes an important aspect of decision-making, meaning that a decision must often be made when relevant data is missing. Under these assumptions the model of bounded rationality was proposed, which implies that (Simon, 1997; Nielsen, 2011):

- Decisions are based on an incomplete and probably inadequate comprehension of the problem to solve.
- Decision-makers will never suceed in generating all possible alternatives.
- The evaluation of alternatives is always incomplete because of the impossibility of predicting all the consequences of their implementation.
- Since it is impossible to determine what alternative is optimal, other criteria are required to make the final decision.

Humans in the position of making a decision in a situation of uncertainty have to rely on judgment and instinct, and use heuristics to reduce complex judgment tasks to simpler ones (Moustakas, 1990). This is a valid approach for most day-to-day situations, as it is usually enough for decision-makers to find a satisfactory solution rather than an optimal one (Nielsen, 2011). In more complex settings where many variables and larger information amount are involved, or when the outcome of a decision may carry significant economical, political, or social consequences, new advances in computer technology allowed research to focus on supporting decision-makers in their task, more notably in organizational and managerial fields in the shape of decision support systems (Eom, 1997; Arnott & Pervan, 2015). Similarly, in personal and consumer decision-making, access to online services can highly leverage the difficulty of finding information to support a choice (Citroen, 2011; Haas et al., 2015), and useful tools exist, such as recommender systems that help to discover relevant alternatives. On the other hand, the interconnected world of today can also accentuate some concerns of the bounded rationality model. Information load and relevance have shown to have effects on decision-making performance (Streufert, 1973); thus, in digital contexts, the enormous amount of data available to decision-makers may at some point be detrimental to their decision-making ability. For instance, achieving a sufficiently informed view of all the relevant alternatives is unlikely, and many may even remain unknown to the decision-maker; or, as another example, having more information does not mean that all is trustworthy or accurate, and contradictions may be encountered. As a result, decision-makers must once again increasingly resort to their judgment and intuition due to the growing complexity of computer-mediated decision environments.

2.1.1 Complex Decision Environments

Previous studies have used the term "complex decision situation/environment" in different ways. For instance, Hogarth & Makridakis (1981) describe a complex environment where a large amount of information can be accessed, and which may change over time due to how several decision-makers interact with it. Kleinmuntz (1985) considers a complex decision environment as one that must be approached from a probabilistic point of view, where decision-makers must base their choice on uncertain cues and their limited information processing capabilities. This can occur when multiple alternatives are potential solutions, but only under certain circumstances. Kleinmuntz also address dynamic decision tasks in which the available information is modified after a choice is made, thus triggering subsequent choices in a feedback loop. Wood et al. (1990) mentions the existence of three aspects that define complexity in managerial tasks: (1) the number of factors that require consideration; (2) the need to coordinate and make trade-offs between different decisions; and (3) the stability of predictive factors in the decision environment. Wood et al. (1990) also states that in a complex decision environment decision-makers must weigh and integrate a wide array of information from diverse sources. Payne et al. (2008) discuss a complex situation where decision-makers must choose between a set of alternatives with many attributes each, and for which relevant information is only briefly accessible and not available at the moment of choosing. Bennet & Bennet (2008) talk about complex situations when referring to those that, among other factors, are difficult to define and change in response to some solution, do not have a single "right"answer, or have many stakeholders. It is also often the case that a single decision cannot solve the problem, but requires a continuing process.

In summary, preceding research mentions "complex decision situation/environment" when referring to either multiple stakeholders, multiple viable alternatives, management of contradictory or incomplete information, integration of multiple sources, and dynamic settings. For the purpose of this research, complex environments are those where either multi-actor or multi-source factors are present, as these type of settings cover most of the aforementioned characteristics of complex decision environments. In multi-actor scenarios it is often required to deal with opposing preferences; multiple alternatives can be valid options viewed from different perspectives; and criteria for choosing alternatives may evolve as group discussion takes place. Multi-source decision environments, on the other hand, are prone to present more uncertainty as there may be contradictory information between sources, or some alternatives may remain unknown if a source is not sufficiently explored; it may be harder to retain and integrate information from different sources; and criteria may change as more alternatives are discovered.

In addition to the multifactorial aspects mentioned above, a technological context also plays a significant role as a further layer of complexity. Although research has shown that information technology is potentially beneficial in the improvement of both efficiency and effectiveness of the decision-making process (Streufert, 1973; Citroen, 2011; Haas et al., 2015), new technological developments may also introduce additional obstacles. In multi-actor decision environments, computer-mediated communication may suppose an increment in decision times, and a decrement in decision effectiveness and quality in comparison to face-to-face meetings (Baltes et al., 2002). Moreover, applying high level of virtuality to tasks of high complexity increases the chances of misunderstandings and mistakes (Marlow et al., 2017; Damian & Zowghi, 2002), and asynchronous communication may also affect the patterns of decision-making and understandings (Berry, 2011). These issues may greatly increase the difficulty of finding a joint decision, or lead to a poorly made one.

As for multi-source settings, information technologies allow simultaneous access to multiple sources (Hilbert & López, 2011), but users are often left alone with the task of understanding, selecting, retaining, and integrating large amounts of information (Wolny & Charoensuksai, 2014). Besides the intrinsic complexities of a multi-source decision scenario, now decision-makers must interpret and evaluate data coming from very different information channels, sometimes as different as a telephonic discussion with a friend, an email, something read on a forum, a digital newspaper, or a physical experience. Without proper support, decision-makers may soon suffer from information overload, a condition where the individual has — or is exposed to, or is provided with — too much information, and cannot make a decision (Levy, 2008; Chen et al., 2009). Furthermore, the phenomenon of bolstering an alternative (Bubnicki, 2013), where decision-makers are biased towards information that reinforces the choice of an option they already prefer and discard information that says otherwise, may be accentuated in situations where contradictory sources are provided.

While multi-actor and multi-source decision environments have received their share of attention, their association with highly technological settings has rarely been studied. Moreover, these situations have mostly been investigated from organizational and managerial points of view, often overlooking that — in today's digitally connected world — multi-factor decision-making takes place in more varied and common scenarios. Therefore, further research is required, first for understanding the challenges that decision-makers must face in modern digital settings, and second to develop appropriate methods to provide support in real world scenarios.

2.2 Multi-Actor Decision Environment: Group Decision-Making

In group decision-making (GDM), individuals collectively decide on an alternative to choose. There are several factors that determine the quality of the decision, such as the decision rules used, available time, or inner group dynamics. In all circumstances, discussion is a key element of GDM, as it is used through the whole process: from explaining and understanding the problem, to defining goals and searching alternatives, as well as establishing the preferences and criteria to evaluate them. Through discussion, participants are able to both present their own ideas and evaluate those of the others, an activity that is beneficial for improving critical thinking and planning skills (Gokhale, 1995; Gauvain & Rogoff, 1989). Due to this interaction between members, group decision-making has advantages and disadvantages over decisions made individually, some of them listed below (Lunenburg, 2010b):

Advantages

- Greater sum total of knowledge: knowledge and experience of some members complements those of the others.
- Greater number of perspectives considered in the decision-making: each member contributes with a unique way of understanding and approaching each decision-making stage.
- Greater number of alternatives: more knowledge, varied decision-making patterns, and group discussion generated by them, may result in the development of alternatives that one member alone would not be able to conceive.
- Increased acceptance of a decision: participating during the process of making a decision increases the possibilities of acceptance, compared to a decision made entirely by others.

- Better comprehension of a problem and decision: participating in the process helps to better understand the logic that lead to the final decision.

Disadvantages

- Social pressure toward conformity: with the purpose of avoiding conflict or being judged by others, group members may comply with a proposed solution even if they do not believe it is adequate, which may result in non-optimal decisions.
- Individual domination: rank, status, and personality play a role in group dynamics, by which equal participation may be hindered. Thus, some members may have the power to make unilateral decisions.
- Undesirable compromises: opposing views can often lead to a compromise or middle ground solution, which may not be the one that provides the greatest overall benefit.
- Time: more time is needed to bring the group together and for the discussion that takes place during each stage of the decision-making process.

As a summary of these points, GDM is useful in tackling complex decisions that require collecting or processing large amounts of information, or where decision acceptance is important for its successful implementation; nevertheless, some negative aspects of group dynamics are also present, most of them related to the way the discussion is moderated, the degree of satisfaction of individual participants, and the adequacy of the final choice. Furthermore, the effectiveness of GDM is largely dependent on the appropriate understanding of the problematic and the requirements for an effective choice, and the appropriate assessment of positive and negative qualities of alternative choices (Hirokawa, 1988).

Another relevant aspect of GDM are decision rules, i.e. the method by which a decision is made. Decision rules can be placed on a continuum depending on who makes the final choice (Sager & Gastil, 1999). According to this, "decision by expert" and "decision by authority" fall at the low (autocratic) end of the continuum, "decision by minority or majority" in the middle, and "decision by consensus" at the high (participatory) end of the spectrum. Majority and consensus rules are most often preferred by people, as they are seen as the fairest options (Johnson & Johnson, 1991). Majority rule is generally more easily applicable, and several simple strategies exist for its implementation (Masthoff, 2004):

- Plurality voting: each decision-maker can vote for his or her most preferred alternative(s), and the one with the most votes is chosen as the group's final decision.

- Utilitarian strategy: decision-makers provide ratings for each alternative that represent their expected satisfaction if that option were to be chosen. Alternatives are then ranked by how much they satisfy the whole group, which can be done either by adding individual ratings, or multiplicating them. The one alternative that obtains the highest value is selected as the group's decision.
- Borda count (Borda, 1784): decision-makers rank alternatives by personal preference to create a list, and score them according to their position in it (the last receives 0 points, the next 1 point, and so on). The final decision is the one that obtains the highest value after adding up all points across lists.
- Copeland rule (Copeland, 1951): decision-makers rank alternatives by personal preference. Then the Copeland index is used to order them, which consists of subtracting the number of times an alternative loses against others to the number of times it beats other alternatives, where ties count as half a point.
- Approval voting: decision-makers vote for as many alternatives as they wish, which results in the election of alternatives that are not strongly disliked.

While a majority-based social choice may be the most straightforward approach, it has been established that consensus produces more satisfying decisions — regardless of their actual effectiveness — (Sager & Gastil, 2006). However, reaching consensus requires a more convoluted and time-consuming process as all members must agree on the decision (Bressen, 2007), which can last indefinitely or be impossible to achieve under limited time conditions. Besides, consensus may only be more effective than the majority rule when supportive communication takes place, which seems to depend on the presence of a certain set of traits in the participants, that is, agreeableness, extraversion, and openness (Sager & Gastil, 2006).

Besides the challenges that time constraints have over consensus-based GDM, time limitations can also deteriorate the overall decision process and result in worse decision-making no matter the decision rules used. The general effects of time pressure can be summarized as follows (Zakay, 1993):

- Information search and processing are reduced.
- Negative information increases in importance.
- Positive attributes of the chosen alternatives are exaggerated.
- Relevant information is overlooked (denied, discounted, or forgotten).
- Information is processed only while there is time remaining.
- Non-compensatory choice strategies become common.
- Judgment and evaluation are more likely to be incorrect.

The importance of group discussion and negotiation can be easily conveyed from the challenges that members of a group must face in order to make a joint decision. It is very important that clear and efficient communication takes place, so that the presented information can be correctly understood and evaluated. A further complication can be extracted from the definition of decision-making provided at the beginning of this chapter, as it stresses the importance of personal preferences in the evaluation of alternatives; thus, with multiple actors, negotiation between the involved parts must be possible, and reducing the disagreement about problem definition, requirements, goals and criteria becomes fundamental (Baker et al., 2001). In this regard, criteria should be able to discriminate among alternatives, support their comparison, include all goals, non-redundant and few in number (Baker et al., 2001). As the number of decision-makers increases, finding agreement on all these moving parts becomes a complicated task, more so under the more difficult to implement consensus decision rules or limited time conditions.

2.2.1 Computer-Mediated Remote Group Decision-Making

As communication technologies advance, their use as a basis for building group discussion and decision-making becomes increasingly relevant. This importance was further boosted in recent times, when the COVID-19 pandemic forced social distancing, and most people's social relationships, both on a personal and business level, moved to a digital plane (Meier et al., 2021; Shufford et al., 2021).

The use of technology to support non-collocated GDM can be advantageous over traditional (face-to-face) one. For instance, it allows enrolling participants in a conversation regardless of their physical location, and reduces travelling costs in terms of time, money, and stress (Orlikowski, 2002; Hinds et al., 2002). However, real-time communication (e.g. via video call) may not always be the best option, as it may involve issues related to communication effectiveness (i.e. receiving messages in a time-efficient manner) or scheduling convenient meeting times for all participants (Berry, 2006). On the other hand, an asynchronous approach to group discussion may not only be better for reconciling schedules, but creates a more flexible environment where partakers have more time to express their opinions and consider those of the others (Cappel & Windsor, 2000). Moreover, no blocking or interruptions occur, and there is no need to compete for air-time, due to the possibility for various threads of thought and concurrent themes to happen at the same time, where participants do not need to wait for their turn to contribute (Berry, 2006; Cappel & Windsor, 2000). Since in asynchronous communication the information is more readily available at all times and threads are easier to re-examine, it is conducive to deep and reflective thinking, which facilitates making better decisions (Jonassen & Kwon, 2001). Lastly, because social, political, or power context cues are less apparent, asynchronous computer-mediated GDM also supports a more equal participation (Berry, 2011).

On the negative side, computer mediated GDM generally requires greater effort to to communicate with other group members (Jonassen & Kwon, 2001). Longer discussion times are common, and consensus is harder to reach in comparison to face-to-face meetings (Cappel & Windsor, 2000). Text-based communication is the usual method to enable asynchronous conversations (Romiszowski & Mason, 2013), but written discussion lacks social context cues such as intonation, facial expression, and gestures that listeners can perceive to fully comprehend the meaning of a message, and whose omission can lead to misunderstandings (Sproull et al., 1991; Vroman & Kovacich, 2002; Marlow et al., 2017). Other general considerations of computer-mediated communication are related to finding appropriate tools to share information, and coordinating and synthesizing contributions from members in and understandable and accessible manner (Berry, 2006).

In conclusion, computer-mediated communication offers a broad set of advantages to group decision-making, such as asynchronous and synchronous communication capacity, high interactivity, or multi-path discussion, and seems to encourage reflective and critical thinking. However, its own shortcomings are added to the complications of group decision-making, which creates a complex decision environment where technology and group dynamics intertwine. To fully benefit from computer-mediated GDM, it is required to provide participants with sufficient flexibility to express themselves and convey their thoughts, find methods to prevent misunderstandings by presenting information in a clear, structured manner, and facilitate tools to conveniently represent and synthesize criteria, alternatives, and their assessment.

2.3 Multi-Source Decision Environment: Purchase Decision-Making

According to the work by Santos & Gonçalves (2021), a retail channel can be defined as *any medium through which consumers can interact with and purchase from sellers*. Thus, smartphones, personal computers or brick-and-mortar stores can be considered retail channels. Touchpoints, on the other hand, refer to *the specific moments in which customer and brand make direct or indirect contact*. Consequently, touchpoints are mediated by channels, because channels allow for touchpoints to happen (Halvorsrud et al., 2016); but these channels do not necessarily need to be directly controlled by firms, as it is the case of word-of-mouth or independent reviews (Baxendale et al., 2015).

Before the digital era, traditional retailing was performed either by directly visiting physical stores, or via catalogue sale by using telephone or mail. Widespread use of the Internet allowed the appearance of e-commerce, and the subsequent expansion of retailing channels and the creation of a plethora of new touchpoints. Over time companies began to offer both physical and digital possibilities rather than focusing on one or the other. This allowed consumers to move across channels more easily, and supported the rise of multi-channel, cross-channel and omni-channel environments. According to the classification given by Beck & Rygl (2015), multi-channel and cross-channel retail refer to shopping environments where more than one channel are available with no or limited integration, and where some interaction between them may occur. However, these channels remain easily distinguishable, unlike in omni-channel retailing, which offers a seamless experience that integrates all of them. The omni-channel consumer experience promises to be the dominant approach in the near future, and it is currently feasible thanks to smartphones and emerging technologies such as augmented reality (Hilken et al., 2018).

Due to these new shopping possibilities, consumers today expect to be able to travel through offline and online touchpoints without hassle, meaning that similar operations must be possible to be performed in either space, and that what happens in one channel must be transparent to the others. Although the benefits of such kind of integration are clear, it also increases the complexity of shopping behaviours (Huré et al., 2017), as, for instance, the practice of showrooming and webrooming, which are becoming commonplace among consumers (Kang, 2018). In showrooming, clients visit a physical store first, which permits viewing products in person and obtaining a direct impression of their qualities, for later concluding the purchase via an online channel. The opposite of showrooming is webrooming, where consumers evaluate products online but make the final purchase in-store. Not only do these behaviours increase the complexity of studying the consumer's decision-making process, but also showcase the relevance of physical and digital channels during the customer journey, and their individual significance to make a purchase decision.

2.3.1 The Customer Journey

The process of making a purchase decision is often referred to as the "customer journey". The concept consists of the different phases consumers go through on their way to making a purchase. Several approaches exist that try to define what these phases are, such as the hierarchy of effects model (Lavidge & Steiner, 1961), the buying behaviour model (Howard & Sheth, 1969), or the popular and frequently cited five-stage consumer decision-making process (Cox et al., 1983; Kotler et al., 2014), which has been generally accepted in the field of consumer behaviour (Blackwell et al., 2006;

Darley et al., 2010). The five stages of this last model represent a series of sequential cognitive phases that lead to the acceptance or rejection of buying a product, as well as the repercussion that the current purchase decision may have in future ones:

Problem recognition: the need or problem that the purchase is supposed to satisfy or solve is detected.

Information search: consumers gather information to learn about how that need can be satisfied, as well as collect data in relation to the (probably several) possible solutions.

Evaluation of alternatives: the different options are evaluated. They are also compared against each other and against the needs of the consumer, filtering down the most suitable ones.

Product choice: after the evaluation of all alternatives, the consumer finally choses the one that may solve the original problem. However, choosing a product does not necessarily mean that the purchase takes place, as the urgent need to solve a problem may dissipate by the time the choice is made, or perhaps not even the most suitable alternative is convincing enough.

Post-purchase evaluation: the consumer evaluates whether the purchase fulfilled its purpose. The outcome of this stage will most likely have an effect on future purchase decisions, as it will count as prior experience during the evaluation of alternatives phase.

Phase length varies depending on the involvement of the consumer in the decision process. While for habitual or emotional products the decision-making process is usually short, in the case of high involvement ones extended problem-solving has been recognized, because consumers are more likely to need to feel connected, search more extensively, and share consumption experiences (Zaichkowsky, 1985; Solomon et al., 2014). Moreover, the availability of several shopping channels can also extend the duration and magnitude of the retail journey.

The five-stage model has been frequently and successfully applied in traditional retailing, and research exists about its suitability in online settings (Darley et al., 2010; Punj, 2012) and multi-channel retailing (Konuş et al., 2008). Nonetheless, some researchers have highlighted the limitations of classic approaches to properly represent the way by which modern consumers face a purchase decision, particularly concerning the linearity of these previous models and its exclusive focus on cognitive drivers. Newer models have appeared to address these issues, as, for instance, the ORCA (Molenaar, 2016) or the model described by Wolny & Charoensuksai (2014), both adapting the buying process to the various channels and touchpoints available thanks to new technologies, with an emphasis on the non-linearity of the decision-making process within a multi-channel retailing context, and where emotional and behavioural drivers are also added to the cognitive ones.

To exemplify these new trends in consumer behaviour modelling, the stages of the consumer journey proposed by Wolny & Charoensuksai (2014) are listed:

Orientation/inspiration/horizon scanning: customers do not have the intention to buy yet, but they scan consciously or unconsciously the marketplace. They obtain information from friends, product displays, magazines, and online sources, such as blogs, reviews or social networks. This is the phase where tailored advertisements based on consumer preferences are more significant.

Information search: customers have the intention to buy, and collect information prior to shopping. They focus on more specialized sources, such as product reviews, ratings, blogs or friends.

Evaluation: the alternatives are narrowed down, for which customers focus on elements such as price, product characteristics, or availability, and it is common to test products in-store. The main channels used at this stage are physical and online stores, as well as friends and social media for confirmation. At this point, it is important for companies to facilitate easy access to the required information (Court et al., 2009).

Purchase: the final product is chosen and the purchase takes place. The used channels are either physical or online stores.

Post-purchase: the experience is shared through word of mouth to friends and/or in social media.

This new view of the metaphorical journey where several, heterogeneous channels intertwine is also acknowledged by Clark (2013), for whom *the customer journey is a description of customer experience in which different touchpoints characterize consumer's interaction with a brand, product or service of interest*. All in all, this way of understanding the customer journey emphasizes how clients access, interact with, and process information during the pre-purchase phase, and how the simultaneous availability of different information sources (i.e. retailing channels or touchpoints) influence the shopping experience and the final purchase decision.

2.3.2 Information Search During the Pre-Purchase Phase

Besides the definition of decision-making that is included at the beginning of this chapter, a second is also widely used: *decision-making is the process of sufficiently reducing uncertainty and doubt about alternatives to allow a reasonable choice to be made from among them* (Harris, 1998). This definition emphasizes the relevance of the search and information gathering process during decision-making. Schmidt & Spreng (1996) defined information search as *the stage of the decision-making process wherein consumers actively collect and integrate information from numerous sources, both internal and external, prior to making a choice*. More recent texts provide similar definitions, as showcased by Blackwell et al. (2006), for whom consumer information search is *the motivated activation of knowledge in memory or acquisition of information from the environment about the potential satisfiers*. In either case, the search for information is made in both internal (i.e. memory) and external sources, of which the latter ones are of relevance in this dissertation.

Researchers have classified external consumer information sources in different manners, but most of them broadly distinguish between the two dimensions of interpersonal (e.g. family, friends or sales personnel) versus impersonal (e.g. advertising), and independent versus seller-dominated sources (Klein & Ford, 2003). More recently, as the search for information via the Internet became more widespread, scholars acknowledged the relevance of online sources (Klein & Ford, 2003; Kim & Ratchford, 2012). This fact, added to the relevance that traditional methods still seem to hold during the purchase decision process (Maity et al., 2014; Wolny & Charoensuksai, 2014), contributed to the creation of a third dimension to represent the dichotomy between offline and online information sources (Klein & Ford, 2003).

The category of offline information sources includes more traditional means of communication, such as advertisements via newspaper, magazine, TV, radio, pamphlets or posters, recommendations and suggestion made by sales personnel, face to face or telephonic word of mouth, expert opinions in specialized TV programs, magazines or in person, and also direct physical contact with products, such as reading characteristics written on the package, or direct product inspection and testing.

Within the category of online information sources one could include websites, online recommendations and advertisements, vendor communication made via online channels, word of mouth from consumer reviews or social networks, expert opinion from blogs or specialized digital platforms, and the use of digital tools such as price trackers, virtual try-on, or comparison sites.

A pre-purchase search is that which occurs when the person has the intention to buy and is hence motivated by an imminent purchase decision (Schmidt & Spreng, 1996). During an external pre-purchase search, the use of smartphones and other Internet-capable devices allow consumers for free online information access and thus gain a better view of available purchasing alternatives and their properties, which supports their evaluation (Church et al., 2007; Taylor, 2016). Consumers who make use of offline and online channels as they suit them when informing themselves about purchase alternatives appear to be "more knowledgeable" and "in control" of the purchase process (Rippé et al., 2015). However, having access to such large amount of information can lead to longer consumer journeys as processing it involves more thought (Hoyer, 1984). Furthermore, situations are more likely to occur where alternatives are not narrowed down after knowing more about them, but more are discovered instead (Lye et al., 2005; Court et al., 2009). Problems of disorientation, confusion and feelings of overwhelm can arise (Mix & Katzberg, 2015; Lu & Gursoy, 2015), and, particularly in multi-channel environments, the mental integration of offline and online information may add a level of complexity, for which remembering and selecting useful data can quickly become a difficult task (Wolny & Charoensuksai, 2014).

Augmented Reality and Information Search in New Retailing Models

Continuing with the topic of channel integration enabled by new technologies, it is necessary to mention the role played by augmented reality (AR) in current research. AR consists of the visualization of 3D digital elements seamlessly integrated with real-world objects in real time (Azuma, 1997). Therefore, with the addition of online capabilities, AR becomes an enabler of omni-channel experiences (Beck & Rygl, 2015; Dacko, 2017; Hilken et al., 2018). Research has shown that mobile AR apps have a positive effect on customer engagement, customer satisfaction, purchase intention, and the overall shopping experience (Pantano, 2014; Dacko, 2017; Poushneh & Vasquez-Parraga, 2017; Bonetti et al., 2019). More importantly, AR technology has been proven to enhance information search at the point of sale and help customers make a buying decision (Spreer & Kallweit, 2014; Chylinski et al., 2014; Ahn et al., 2015). For all these reasons AR has found its place within the retailing sphere as a viable solution to provide easy access to on-site information during the pre-purchase phase, supported by the contextual awareness and information readiness enabled by the technology. However, AR research is still immature in many aspects, and although examples exist that showcase its capability to enable the access to extended product information (Välkkynen et al., 2011; Gutiérrez et al., 2019), in-store navigation (Cruz et al., 2019), and functionality explanations (Ludwig et al., 2020), there is still a long road ahead to clearly define which interaction and visualization methods better support information search during the pre-purchase phase, and are more suitable for use in physical retail contexts.

2.4 Recommender Systems

Recommender systems (RS) arise from the need to provide users with suitable alternatives from among a large number of them, and for which a complete evaluation of all possible options is unfeasible or would require great effort. RS are conceived to support users in various decision-making processes by analysing, reducing and conveniently presenting information that is tailored to their needs (Ricci et al., 2011), hence alleviating issues such as information overload (Chen et al., 2009).

To achieve their purpose, RS generate predictions and recommendations of items by integrating information from different sources, and balancing factors like accuracy, novelty, dispersity and stability (Bobadilla et al., 2013). The recommendation problem can be reduced to a matter of estimating ratings for items not yet assessed by the user, and then selecting those top N items that maximize the expected outcomes (Adomavicius & Tuzhilin, 2005; Ricci et al., 2011). RS are usually classified by the method by which these ratings are predicted, and literature generally recognizes five of them: demographic, content-based, collaborative, knowledge-based, and hybrid filtering techniques (Burke, 2000; Adomavicius & Tuzhilin, 2005; Candillier et al., 2007; Ricci et al., 2011; Bobadilla et al., 2013; Kunaver & Požrl, 2017).

Content-based filtering

Content-based systems generate recommendations by comparing candidate alternatives against the attributes of items for which the user already gave an assessment in the past. Therefore, items with features similar to other items that were previously highly rated by the user also receive higher predicted ratings, and thus, are more likely to be recommended. More specifically, a user profile is learned by collecting user preferences on the properties of items that the user likes and dislikes; then, this user preference model is used to determine the recommendation score of any unrated item. The user profile becomes more accurate over time, as it is updated when new user interests are observed.

Some limitations of content-based systems are:

- Limited content analysis: the system can only rely on the features contained by those items that the user has already rated. That means that sufficient attributes have to be defined for each item manually, or methods have to be implemented for the system to extract them itself (e.g. from a written description), but information retrieval techniques may not be as accurate with more complex types of data (audio, images or video files).
- Overspecialization: recommendations are limited to those that are similar to items the user already rated, and in some cases even items that are too similar are filtered out to avoid redundancy (e.g. news feed). As a consequence, the resulting set of recommendations may lack diversity, which prevents users from discovering new items that they may also find interesting, but that are different from anything they rated before.

- New user (cold-start): for a user profile to be reliable, a sufficient number of items must be rated first to allow a content-based system to learn from them. Hence, new users may obtain less accurate recommendations (which is known as the cold-start issue).

Collaborative filtering

In a collaborative RS, user-specific rating predictions are obtained by exploiting similarities between users or items. There are two main approaches to collaborative filtering: user-based and item-based. User-based approaches explore similarities between users, i.e. their rating patterns: missing ratings on an item for a particular user are filled by extrapolating the ratings given to the same item by the user's nearest neighbours. Item-based approaches, on the other hand, predict ratings based on item similarity: the assigned rating to a given item on a given user is calculated by analysing the ratings that the user provided for other similar items.

Collaborative RS avoid some issues encountered when using content-based filtering techniques. Because they provide recommendations based solely on ratings, the type of content that is being recommended has no impact on the process. Furthermore, user-based approaches are capable of recommending items unrelated to what the user has rated so far, but that other similar users did, hence offering more diverse recommendations. However, other issues may arise instead:

- New user (cold-start): as with content-based filtering techniques, the system requires first to learn a user's preferences in order to generate accurate recommendations.
- New item: items with no ratings cannot be included in the recommendations because the system lacks any information about them. This is the case for recently added items, in which case time is needed until sufficient ratings are obtained.
- Sparsity: generally, the number of ratings obtained are very small compared to the number of unrated items, and to be able to generate recommendations effectively, an RS requires to reach a critical mass of users. Not fulfilling this requirement can lead to poor recommendations due to a lack of users similar to the current one (known as the grey sheep problem), or because relevant items do not have enough ratings yet.

Knowledge-based filtering

A knowledge-based RS makes use of predefined rules or "knowledge" about items on a certain domain to generate recommendations, in order to meet a set of requirements provided by the user. It offers advantages over previous filtering techniques, since no user base is required and gathering user information is typically done in a conversational manner (thus, the cold-start problem is not present). However, adding new features requires knowledge-engineering and, without a learning component, the ability of the system for providing suggestions remains static and does not improve over time, unlike collaborative and content-based filtering techniques.

Hybrid filtering

Hybrid approaches combine collaborative and content-based filtering techniques, with the goal of overcoming their individual limitations and reaching peak performance. The different ways by which this is achieved can be summarized as follows:

- Implementing each approach separately and aggregating their outcomes.
- Incorporating some characteristics of one approach into a system that fully implements another.
- Implementing a completely unified model.

Hybrid approaches can be further extended by including knowledge-based techniques to address the cold-start issue and improve the general recommendation accuracy.

In addition to sparsity, overspecialization and cold-start issues, RS also have to face more general problems like scalability (performance of the system is often reduced as the number of users and items increase), robustness against untruthful data (e.g. fake ratings inserted to decrease the recommender's accuracy or influence the recommendation outcome), and providing recommendations sets that include diverse and novel items (Khusro et al., 2016; Kunaver & Požrl, 2017).

Regardless of how recommendations are generated, the more information the system has of users the more elaborated and accurate recommendations can be. Although some information can be retrieved implicitly — e.g. browsing history, clicking behaviour, location, social information (Kelly & Teevan, 2003) —, it is often required to directly address users or let them have more control over their preferences. This is achieved via interaction, which may occur at two stages of the recommending process: user preference elicitation, and result presentation (Jugovac & Jannach, 2017). During the preference elicitation phase the system uses mechanisms to acquire explicit information about user interests, which are to be utilized to generate recommendations. Some examples are ratings, likes, static profile forms, conversational interfaces, critiquing (i.e. user assessment of the features of a proposed item), or quizzes (to infer information the user may not be aware of, such as personality traits or buying behaviours). After acquiring an initial set of preferences and filtering matching items, these are presented to the user. At this stage, the system may allow users to give feedback and refine their preferences. This requires for users to be able to explore and learn about the suggested recommendations, thus obtaining a better idea of their qualities and suitability. Visual presentation of items must be taken into consideration, and offering access to relevant data and explanations is crucial (Papadimitriou et al., 2012). Users can then modify their original preferences accordingly, or evaluate recommendations via rating or critiquing, which usually triggers a new iteration of the recommending process that can be repeated until one of the proposed alternatives is chosen.

This introduction to RS has so far presented an overview of a number of challenges that are assiduously addressed in past and current research in the field. Some of these difficulties can be intensified when applying RS to complex multi-factor scenarios like the ones studied here. For instance, a multi-actor decision environment requires for recommended items to fulfil the preferences of several users at once. Questions arise about how individual preferences can be defined and aggregated, or how recommended items are assessed and selected, all of them issues covered by a subfield of RS know as group recommender systems (Felfernig et al., 2018). Similarly, the multi-source scenario has gained relevance in RS research due to the introduction of the Internet in most aspects of daily life, which has fostered the use of recommendations in a variety of areas (Park et al., 2012; Kunaver & Požrl, 2017), hence creating new opportunities for RS to access and incorporate different types of information in their recommendations (e.g. product data, user-created, geo-social, or knowledge-based). However, this has forced RS to use hybrid architectures through the employment of different technologies, each suitable for specific types of information sources (Bobadilla et al., 2013), and new challenges have emerged in the search for appropriate methods to integrate and present information to the user.

2.4.1 Group Recommender Systems

Group recommender systems (GRS) facilitate joint recommendations to groups of users by considering their individual preferences (Felfernig et al., 2018). Since their first implementation by McCarthy & Anagnost (1998), GRS have grown in relevance and have been exploited in different areas involving group activities, such as listening to music (Crossen et al., 2002; Chao et al., 2005), watching television (Masthoff, 2004; Yu et al., 2006; Quijano-Sanchez et al., 2011), travelling (Ardissono et al., 2003; Amer-Yahia et al., 2019), or going to a restaurant (Park et al., 2008).

GRS usually have a first stage where individual user information is collected. Afterwards, a collaborative filtering approach allows for four different techniques by which group recommendations can be generated depending on what and when user information is aggregated (Bobadilla et al., 2013; Ortega et al., 2013):

Aggregation of recommendations: it occurs at the final stage of the recommendation process. Recommendations for each user are calculated separately, and the set for the group is made by merging them.

Aggregation of predictions: it happens after calculating rating predictions for each user. These predictions are aggregated and then used to generate a recommendation set.

Aggregation of neighbours: neighbours for each user are found based on their individual similarity scores, and then joined to create a single neighbourhood of users for the whole group.

Aggregation of preferences: individual user preferences are merged to create the group's preference model, which is directly used by the similarity metric.

Although the quality of the recommendations remains similar among approaches, research shows that execution times are reduced the earlier in the process the aggregation takes place, and therefore preference aggregation is the most efficient technique (Ortega et al., 2013). There is a wide range of methods by which the group's preference model can be constructed: from aggregating all individual preference models (Lieberman et al., 1998; McCarthy & Anagnost, 1998), to creating homogeneous subgroups and finding a model that satisfies them (Ardissono et al., 2003), or including additional group variables that are added to individual preferences (Beckmann & Gross, 2010). Independently of the composition of the data to be combined, the aggregation function by which user information is translated into the final group preference model also differs between existent GRS (Masthoff, 2004). The most straightforward method is the average strategy, where the average rating of all members on an item is used as its group score (Ardissono et al., 2003; Jameson, 2004); the least misery strategy focus on maximizing overall satisfaction by scoring items based on their minimal individual rating (O'connor et al., 2001; Beckmann & Gross, 2010); the average without misery strategy is a mix of the previous ones, which consists of using the average score but discarding those preferences where individual ratings are below a lower limit (McCarthy & Anagnost, 1998; McCarthy et al., 2006); finally, another commonly used method is the median strategy, which uses the middle value of the rating of all group members (Jameson, 2004).

Nonetheless, all these approaches rely on the pre-existence of well-defined user preference models, which is a requirement hard to meet under certain conditions, such as for occasional groups that gather spontaneously or when user data is distributed among different systems. A further problem arises from the variability in user preferences due to situational context; that is, users may define a set of preferences when in a group situation that differs from what they would normally choose, for the things enjoyed alone are not necessarily the same as those enjoyed with others. The most direct way of addressing these limitations is to directly ask users to define their individual preference model prior to the recommending process, as it is the case for AGReMo (Beckmann & Gross, 2010); a different approach is used in the Travel Decision Forum (Jameson, 2004), where users always start with an empty profile, and collaboratively agree on specific group preferences during each session; on the question of the variability of preferences depending on the context, in the Collaborative Advisory Travel System — CATS (McCarthy et al., 2006) — users can incrementally modify individual preferences to adapt them to the specific needs of the group and influence the outcome of the recommendations. However, even in these instances, group interaction during the preference elicitation phase is barely supported, and instead occurs mainly at late stages of the recommendation process (when members must choose an alternative among the recommended items), leaving users with little control over the process itself. This is contrary to what happens in real world situations, where multiple actors facing group decisions interact from the beginning by expressing their preferences, reviewing those of others, and revising their own, until reaching a joint conclusion (Baker et al., 2001), and research has already highlighted the important role that group discussion also plays in reaching consensus within the GRS field (Basu Roy et al., 2010). Additionally, preventing users from actively participating in the construction of the group's preference model may also have repercussions on the acceptance of the proposed recommendations as it can be harder to understand the process that led to them, and therefore making a final decision may pose a greater challenge (Lunenburg, 2010b).

Muti-actor decision-making is mainly concerned with collaboratively making choices from a social point of view, where interaction, discussion and negotiation are key factors. In that regard, research on GRS is increasingly concerned with supporting group decision-making mechanisms through the addition of interaction-based and social functions (Alvarado Rodriguez et al., 2022). Initial approaches had to rely on face-to-face communication, as it was the case for the system proposed by McCarthy et al. (2006), in which a collocated group could discuss recommendations and each other's preferences around a multitouch table. However, in today's world group discussion is more often performed thanks to computer-supported communication, reason for which more modern approaches to GRS have appeared that include elements such as discussion chats, or cues to indicate user intentions or reactions, to facilitate recommendations for non-collocated groups (Nguyen & Ricci, 2018).

Nonetheless, a research gap exists with respect to the full integration of a GRS within a multi-actor decision setting, which should be able to support group negotiation and discussion at all stages of the recommendation process, from preference elicitation to the selection of an acceptable alternative. User interaction comes into play early and at various points: first, if users are aware of individual preferences of other members, they may as well try to convince them to make modifications; if users are granted direct control over what preferences are added to the group's model, negotiation is necessary to decide their inclusion and importance; and finally, selecting an alternative requires for users to discuss its suitability, and probably make concessions and revisit group and individual preferences to reach mutual agreement. So far, no research exists on a system that allows control, and provides discussion and negotiation tools, over personal constraints, group preferences and the final selection of alternatives (Alvarado Rodriguez et al., 2022). Hence, there are open questions on how to effectively support decision-makers during the formulation and negotiation of preferences, as well as to reach group consensus. In addition, it is also necessary to define appropriate methods for presenting and organizing information on individual and shared spaces that can be influenced by several actors at the same time.

2.4.2 Recommender Systems in Physical Shopping Contexts

So far, it has become clear the relevance of the Internet and its increasing ubiquity, firstly as a source of complexity in multi-factor decision environments, and secondly as the enabler of possible solutions. When it comes to retailing, the importance of the Internet has to be stressed once again, as it made possible the rise of online stores, and the more recent trends related to offering online services within traditional retailing settings (Beck & Rygl, 2015).

Online shopping generally entails greater access to alternatives and information, and therefore greater difficulty in making decisions (Perea y Monsuwé et al., 2004; Chen et al., 2009). This has led to the development of mechanisms to support the purchase decision-making process, such as comparison tools, price trackers, customer reviews and ratings, or detailed product descriptions (Kocas, 2002; Park & Gretzel, 2010; Lackermair et al., 2013). With the same purpose, RS have also been successfully applied into online shopping environments, where they can collect implicit and explicit user information, as well as access several information sources to provide more accurate product recommendations (Schafer et al., 2001). Research on RS applied to e-commerce is extensive (Ricci et al., 2011), and has proven their usefulness during the purchase decision-making process to speed up the process of product filtering, discovery, and information seeking, as well as their capacity to transform browsing-only clients into buyers, improve consumer loyalty and provide cross-selling opportunities (Schafer et al., 2001; Kourouthanassis et al., 2002). It is therefore not surprising that many popular commerce websites have already included RS for some time now, as it is the case for Amazon or eBay (Schafer et al., 2001).

More recently, the growing availability of online-capable and context-aware mobile devices have fostered the investigation of RS in physical environments too, which, in addition to many other applications, can be used to support customers by providing in-store recommendations (Abbar et al., 2009; Ricci et al., 2011). RS may be particularly useful in brick-and-mortar stores, where clients usually lack well-defined preferences at the moment of their visit, and, instead, tend to build them progressively as they learn about the product space (Payne et al., 1992). Existing research that explores this topic is mostly focused on offering information and recommendations based on the consumer's location, surroundings, or behaviour. For instance, the system presented by Kourouthanassis et al. (2002) keeps track of the clients shopping list and offers promotions based on buying behaviour; Sae-Ueng et al. (2008) developed a RS that provides recommendations based on observed interactions towards the available physical products; von Reischach et al. (2009) introduced a system for mobile devices that lets consumers receive product data, recommendations and user ratings of products scanned at the point of sale; Chen et al. (2015) investigated a smart environment where RFID is used to collect user contexts with which to generate product recommendations; or the research by Fagerstrøm et al. (2020), which explores the benefits of using the Internet of Things (IoT) to create personalized offers based on products in the basket. Prior studies demonstrate that using RS in physical environments offers advantages such as obtaining a better knowledge of the product space and reaching more informed purchase decisions. However, most research focuses on displaying recommendations and other information on static or handheld devices, which means that real world offerings, and digital information and recommendations are presented on different spaces, which also differ in the way clients interact and navigate through them. A more homogeneous and integrated visualization of product data and recommended items may grant benefits in terms of information acquisition and comparison, and thus better support purchasing decisions.

Augmented Reality as Provider of In-Store Recommendations

As it was already highlighted at the end of Section 2.3.2, AR is particularly well suited to provide on-site digital information that blends with physically present items, while also bringing advantages in terms of information search and user experience. Therefore, enabling recommendations in physical shopping contexts through digital augmentations would allow the combination of the enhanced decision support afforded by RS with the engaging experience delivered by AR. This represents a promising step forward in both RS and AR research fields, and there are already indications that AR-enhanced product recommendations may offer some benefits in contrast to standard browser based UIs (Huynh et al., 2018). The use of AR also opens the door to investigating new retailing models, where digital catalogues could be explored through the recommendations given for physical ones, and 3D augmentations would enable their comparison.

Scientific and commercial spheres have already shown interest in exploring AR-based solutions to support consumers, mostly in the shape of virtual try-on (Kim & Forsythe, 2008; Javornik et al., 2016; Smink et al., 2019), or mobile-apps like the Skin Advisor by Olay (2020), which detects the user's face skin condition and recommends suitable products. In spite of it, the issue of providing AR-based in-store recommendations has been largely overlooked, and only few examples exist in the literature (Zimmermann et al., 2022). To name some, Aisle 411 & Tango (2014) partnered to create an AR-based app for Walgreens stores to deliver product information, personalized promotions and navigation; the research by Torres-Ruiz et al. (2020) where the IoT and AR are used to recommend itineraries in a museum, although not focused on retailing, could be applied to physical commerce; Mora et al. (2020) discussed the use of mixed reality to implement in-store shopping assistants that offer personalized product recommendations; recently, Zimmermann et al. (2022) developed a prototype app for smartphones capable of providing in-store assistance including explainable recommendations, the study of which indicates that AR assistance offers benefits in terms of usefulness, entertainment and informativeness in contrast to unassisted shopping, but also result in a more "irritating" process.

Altogether, there is sufficient indication that AR can be used to support the purchase decision-making process by offering in-store recommendations, and other services such as contextualized product data, explanations, navigation and easier comparison of alternatives. Still, due to the novelty of the approach, there are many open questions concerning how these functions could be implemented to better support customers, specially considering the idiosyncrasy of the consumer-journey in modern multi-channel shopping contexts. Moreover, most consumers have little or no experience with AR, less so when it is used for non-ludic purposes. It is thus unclear what visualization and interaction techniques are more appropriate, how consumers will navigate and learn in a hybrid digital-physical shopping environment, or what impact it may have on their decisions. There is also a need for investigating the driving factors in the acceptance of AR-based shopping assistance; classical theories such as the technology acceptance model (Lee et al., 2003) may not suffice to justify acceptance in modern retailing contexts, where other factors, such as personality traits (Zimmermann et al., 2022; Hermes et al., 2022), may be of greater relevance.

On the basis of what has been presented in this chapter, recommender systems appear to offer the right tools to support decision-making in the multi-factor scenarios studied in this dissertation: their main purpose is to support decision-makers by filtering and homogenizing information from multiple sources and actors; they have proven efficacy and extensive research in group and shopping environments; and they are widely used and constantly evolve to make use of, and integrate with, modern technological currents. Nonetheless, there are still significant research gaps regarding their successful application in both multi-factor scenarios.

In the case of remote multi-actor decision environments, it is still required to find appropriate tools to create and modify group preferences on the fly, while giving users control over what information is shared with others. Due to the relevance of interaction, discussion and negotiation between members in group decision-making, these factors must also be present during the whole recommending process; however, such elements have been largely under-explored in the GRS field, and it is still unclear what information-sharing and consensus-finding methods that support social interaction may be more suitable under a remote multi-actor decision context.

On the other hand, the use of RS in multi-source decision environments within a physical retailing context promises to bring benefits during the purchase decision-making process, as suggested by previous research. However, in recent years, there has been a revolution regarding how user data is collected and how users access and interact with information (e.g. the emergence of the IoT or AR), for which research is still scarce in shopping scenarios. AR appears to be particularly well-suited to provide assistance during the purchase process, and brings exciting new opportunities to present and interact with digital recommendations. Additionally, its ability to blend digital and physical information further enhances and complements that of RS as source homogenizer. Nonetheless, many questions remain unanswered regarding the possible implications that this new shopping paradigm may have on the way customers learn, explore, and evaluate alternatives, and make decisions.

3 Research Goals and Application Scenarios

3.1 Research Goals

> **Research Goal 1**
>
> Identify and describe real-world complex multi-actor and multi-source decision environments and investigate the challenges that they present.

The first research goal of this dissertation is to identify specific scenarios where complex multi-factor decisions take place, with the purpose of better understanding the implications of these type of settings and the challenges that decision-makers in these situations have to face. Identifying all the intervening elements, and investigating real world scenarios and their limitations, may be determinant for developing strategies to provide decision support in complex multi-factor environments.

Based on the description of complex decision environments given in the background section (2.1.1), the multi-factor decision scenarios considered here can occur at two levels: multiple actors and multiple sources. Therefore, this research is focused on two different application settings, that is, a situation where multiple actors are involved (group decision-making), and another where only one actor must make a decision, but several sources of information are available (purchase decision-making). In addition, the (limiting) role that technology plays in these contexts is also taken into account, as it adds another level of complexity. A more detailed description of each application scenario and its specific research questions is provided later in this chapter.

> **Research Goal 2**
>
> Develop and evaluate suitable methods by which recommender systems can be used to support decision-making in complex multi-actor and multi-source decision environments.

Through the investigation of the aforementioned scenarios and the challenges they pose, this research aims at designing and developing suitable methods to support the decision-making process that takes place in them.

Multi-actor and multi-source settings, while similar in many aspects, differ in crucial elements that makes them unique, and for which they require solutions adapted to their particular idiosyncrasy. Consequently, it is within the scope of this dissertation to design approaches that offer decision support with a focus on the two defined multi-actor and multi-source decision environments, and that make use of current technological advances and recommending techniques for their implementation.

A further goal consists in the assessment of the developed approaches, with the purpose of drawing conclusions about their feasibility and effectiveness in relation to specific criteria relative to each scenario.

3.2 Application Scenarios

3.2.1 Preference Negotiation and Decision-Making in Group Recommender Systems

This first application context represents a situation where multiple actors must collaborate in defining of the preferences used to generate group recommendations, and jointly make a final choice among the given alternatives.

Usually, group decision-making allows for more informed and better decisions, because it signifies the sum total of knowledge and information that each member individually possesses, as well as more variety regarding possible approaches to solve a problem (Lunenburg, 2010b). However, it greatly relies on social factors related to discussion and negotiation, through which the group can cope with disagreement between members, and the different alternatives can be evaluated. Making a final choice on a matter can be accomplished in different manners, such as unilaterally (when only one group member decides), by majority, or by consensus (Bressen, 2007). The latter is commonly seen in daily situations, like friends deciding on a restaurant or a movie. However, reaching consensus is often the most problematic method, as it means that all group members must agree with the chosen solution. It also implies that members must be able to express their own preferences and sufficiently understand those of the others, which requires a certain set of personal skills and group dynamics to take place (Sager & Gastil, 2006).

Traditionally, group decision-making occurs within a physical space, where all participants can discuss and present their arguments and preferences in person. When members of a group are face-to-face, the one presenting his/her point of view has the possibility to support the explanation on body-language (e.g. voice tone, look, or hand gestures) that may help in conveying the message to the other members; likewise, feedback can be more directly received. Nonetheless, issues such as social pressure or individual dominance may also arise (Lunenburg, 2010b).

The rapid development of telecommunications technology in the last two decades has changed how people communicate (Romiszowski & Mason, 2013). This is also true for how groups meet, and thus, the space where they make decisions (Berry, 2006). A virtual setting offers advantages over a physical one, as some of the negative implications of physical group meetings are mitigated — e.g. by allowing a more equal contribution from all participants (Berry, 2011). Nonetheless, discussion is heavily constrained by the limitations of the medium through which the communication takes place, and expressing oneself in a precise manner is more complicated (Marlow et al., 2017; Damian & Zowghi, 2002). Furthermore, group discussion supported by digital technology is commonly asynchronous, which, although helpful in reconciling participant schedules, also means longer and interrupted decision times (Cappel & Windsor, 2000).

In addition to the aforementioned challenges of remote group decision, it is required to acknowledge the complications deriving from a consensus type of decision-making and the complexity of finding a solution acceptable by all group members (Bressen, 2007). Due to the digital context considered here, group recommender systems are a viable approach to help achieve group consensus, as they facilitate joint recommendations based on the group's preference model (Felfernig et al., 2018). Recommended items can then be taken as suitable alternatives that reduce the decision space. However, group preferences are often the result of group interaction, and tend to be created on the fly either because they depend on the situational context or the group gather spontaneously. These considerations demand ways by which members of a group can express their individual preferences and influence the outcome of the recommendations dynamically, as well as methods to support discussion and negotiation of group preferences. Nonetheless, existing approaches to GRS typically only consider group interaction after the recommendations have been presented, and offer little or no agency in the creation or direct manipulation of group preferences (Alvarado Rodriguez et al., 2022).

Altogether, remote group decision-making with negotiation of preferences and consensus support makes for a complex decision environment (Figure 3.1) that has not been sufficiently researched, but that represents a fairly common scenario. To further investigate it, a concept was designed for a platform that supports all stages of the group decision process, facilitated by a group recommender system. Non-collocated members of a group collaborate during the preference elicitation stage for creating a shared preference model, which will be then utilized for generating group recommendations. Members can discuss and negotiate about individual and shared preferences, which can be recursively modified through group interaction, until a suitable recommendation is found. Several prototype systems were developed along these lines in an incremental process, all of them focused on the hotel domain, with the aim to allow members of a group to jointly decide on which one to book.

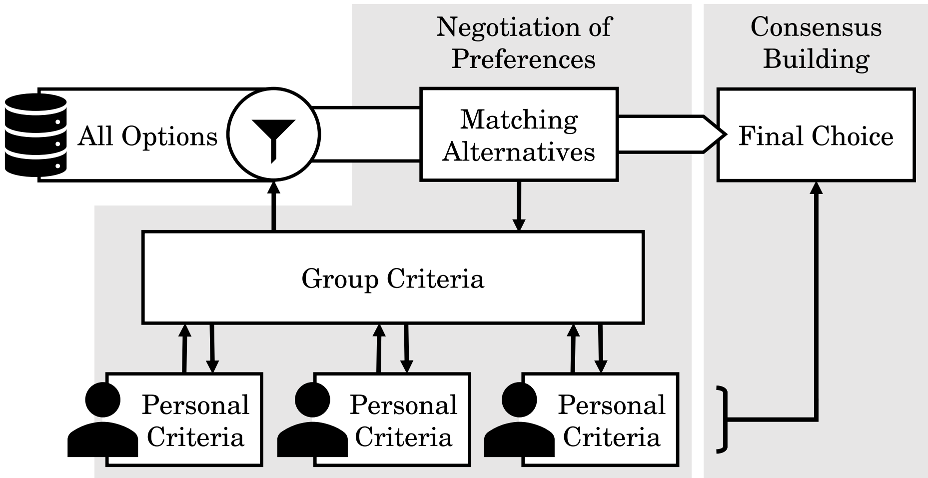


Figure 3.1: Decision-making process of a multi-actor decision scenario with negotiation of preferences during the group's preference elicitation stage, and consensus building to make a final decision. Negotiation may occur at group and personal levels, after the evaluation of current criteria and matching alternatives.

Research Questions

> **RQ1** What are effective means for supporting the formulation, the exchange and the negotiation of group preferences in multi-actor recommending and decision-making?

> **RQ2** How to structure the preference elicitation and construction process into private spaces for setting up one's individual preferences versus public spaces that can be seen and criticized by the whole group?

> **RQ3** How do the proposed approach and developed system perform in terms of usability and acceptance?

3.2.2 Supporting Purchase Decisions in a Hybrid Physical-Digital Setting

The application scenario focuses on a setting where relevant information is scattered across different heterogeneous sources, and for which joint access for data collection and examination is not trivial.

Clients of a traditional physical store, may often make a purchase decision based on a limited set of options and little available information. First, they can only choose from among those items that are present at the store, which in many cases greatly reduces the number of alternatives. Second, the information that they can obtain of each available product can only be acquired from nearby sheets or posters, the seller's advice, or direct physical inspection of items. This creates a scenario with a great deal of uncertainty, either regarding the existence of other possibly better options that are not available at the store, or concerning the completeness and veracity of the information about the products that are physically available, all of which may prevent consumers from making a confident purchase decision. For this reason, it is often the case for clients to make their own research before actually visiting the store, and take time to learn more about the product space by, for instance, identifying the most important product characteristics, finding popular options or brands, asking opinions from friends or family, or requesting expert advice (Wolny & Charoensuksai, 2014). In this manner, it is possible to make a more informed purchase decision even within the constraints of physical retailing, an issue that takes on additional importance when dealing with high-involvement products, for which making the wrong choice can have greater and longer-lasting consequences (Zaichkowsky, 1985).

Since the apparition of the Internet and the further development of communication technologies, the way by which consumers learn about products and evaluate their alternatives has experienced a change, and now more possibilities exist to collect relevant product information during the pre-purchase phase (Klein & Ford, 2003; Wolny & Charoensuksai, 2014). Not only product characteristics are usually easily accessible through spatialized websites, but also the opinions of other buyers, and extensive reviews made by experts. The readiness and accessibility of this information, by itself, supports making a better purchase decision and finding an adequate product (Rippé et al., 2015); however, it also means that consumers have access to a broader set of possible purchase alternatives, whose evaluation and comparison may result in greater mental effort and choice overload (Chen et al., 2009). As a consequence, narrowing down an acceptable product may become a more complicated and frustrating task, and multiple approaches have been taken to alleviate the challenge of filtering and evaluating all the available options, such as product comparison tools and recommender systems (Schafer et al., 2001; Park & Gretzel, 2010). Nevertheless, after deciding on a product by using online methods, many of the final purchases are completed in physical stores, in a process that is known as webrooming (Kang, 2018), and which highlights the importance that physical retailing still has for many consumers (even among those who make frequent use of online services).

Modern technology allows a new scenario to take place, where users of online shopping functions have access to them within the context of physical retailing, mostly thanks to the use of smartphones and other mobile devices (Taylor, 2016). In such situation, clients have simultaneous access to several information sources from either online or physical domains. The nature, accessibility, presentation, and interaction with these sources may vary greatly from one to another: from reading physical plantlets, or direct product inspection and talk with sales personnel, to visiting online portals, reading product reviews, using online shopping tools, or discovering alternatives not present in the shop's current selection of products. Despite the many benefits that having access to more complete and varied information has during the purchase decision process, it may also bring along issues that were already present in online-only environments, in relation with human capacity to assess all possible alternatives and process all available information (Chen et al., 2009). Additionally, it may as well create new problems, such as the challenge of mentally associating digital information with that which pertains to the physical space, all of which creates a unique and rather complex decision environment. Although the seamless integration of retailing spaces is currently a trending research topic under the term of omnichannel retailing (Beck & Rygl, 2015), not enough research has yet been done from a decision support perspective to cover its feasibility, as well as other aspects such as user acceptance of these type of systems or the implications that mixed online-physical settings may have on the purchase decision.

To answer these questions, a new approach to shopping support systems was designed and implemented for augmented reality head-mounted displays (Figure 3.2). The system makes use of new AR advances to display digital data anchored to real-world products. These augmentations provide detailed information about product properties, including explanations and 3D visualizations of their components to better understand their characteristics. Product recommendations based on physically available items are generated and displayed next to them to support the discovery and exploration of new ones. These recommendations can be refined by critiquing the attributes of the products they are based on. Furthermore, recommended items may contain products not physically available, but that are part of the store's digital catalogue, which integrates and expands the accessibility to more possible choice alternatives. Lastly, the system features a comparison tool that allows users to compare the attributes of up to three products at once, no matter to which reality domain they belong. Several prototypes were developed and evaluated, each including new incremental features, and focused on vacuum cleaners as product domain.

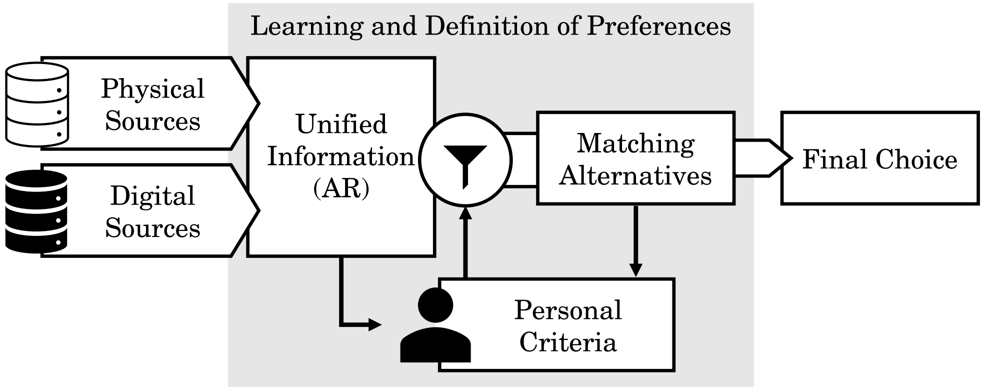


Figure 3.2: Decision-making process of a multi-source decision scenario with integration of online and physical sources via AR. Decision-makers build their preferences by exploring the unified information and evaluating current matching alternatives.

Research Questions

> **RQ4** How can AR-based functions in a physical shopping scenario support the decision-making process in relation to information access, product discovery and purchase confidence?

> **RQ5** What are the implications that a hybrid physical-digital setting may have on the way users learn and explore the digital space?

> **RQ6** What is the significance of consumer psychological factors in the adoption of AR-based shopping support systems?

4 Contributions of the Papers Included in this Dissertation

Due to the cumulative nature of this dissertation, the outcome of the research presented here is grounded on the results obtained in a series of related papers. Each one of these papers addresses one or more research questions of those specified in section 3.2. However, it may be the case for a paper to not entirely answer a certain research question, which means that no conclusive results were obtained, or only some particular aspects were covered, in which case further research was required. Questions that were answered in one paper were taken as assumptions in subsequent ones, allowing the overall research to progress. Nevertheless, papers are self-contained and can be comprehended entirely on their own.

This section provides a summary of all the papers that contributed to this dissertation, the complete version of which can be found in the Appendix. For each paper, the following information is provided: title, authors, publication venue, current publication status, and a summary of its background, approach and results. Additionally, research questions addressed by each paper are indicated, where ● signifies that it was fully addressed, and ◐ that only partially.

4.1 Paper I

Table 4.1: Summary of Paper I. Full manuscript on page 75.

	Preference Elicitation and Negotiation in a Group Recommender System
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	15th IFIP TC 13 International Conference on Human-Computer Interaction – INTERACT 2015
Type	Conference Paper
Status	Published
RQs	RQ1: ◐ RQ2: ◐ RQ3: ◐
Background	Group recommending presents challenges such as the collection of information from separate individual users, the situational variability of group members' preferences, or solving the complex balance between individual and group satisfaction. These aspects are closely related to group decision making, particularly the social interaction that occurs when a group of people must make a joint decision, for which they collaboratively define criteria and choose an alternative solution. In spite of that, current group recommending techniques generally overlook the significance of group interaction.
Approach	To address this research gap, a novel approach to GRS was designed that better takes into account group interaction during the recommending process. First, group discussion is supported during the preference elicitation stage by allowing users the definition and negotiation of their own individual preferences. These process consists of the definition of private features that can be added into public lists of desired attributes ordered by importance (i.e. the user's preference model). Individual preferences are aggregated in real time to generate the group model, which is also made available for direct inspection by the users. Second, the evaluation of alternatives is assisted by providing tools that facilitate consensus building to collectively select a final choice. These ideas were implemented into a GRS prototype (Hootle) that generates hotel recommendations for a non-collocated group of people. The system was evaluated in a user study (48 participants) to investigate the performance of the approach regarding different group sizes and in comparison to a baseline GRS without negotiation support.
Results	Results indicate that including discussion support during the recommending process increases the overall satisfaction and perceived recommendation quality. However, these effects are more noticeable in small groups (3 members) and are diluted the larger the number of participants, which calls for further investigation of the effects of group size and optimization of the scalability of the approach. Related to this, participants expressed concern about the complexity of tracking changes in individual user profiles, which makes their negotiation more difficult and results in larger group preference models.

4.2 Paper II

Table 4.2: Summary of Paper II. Full manuscript on page 95.

	Hootle+: A Group Recommender System Supporting Preference Negotiation
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	22nd International Conference on Collaboration and Technology – CRIWG 2016
Type	Conference Paper
Status	Published
RQs	RQ1: ◐ RQ2: ◐ RQ3: ◐
Background	An approach to GRS with discussion support to allow the negotiation of user preferences and group consensus building was evaluated in previous research (Paper I). Results showed that despite providing advantages in terms of user satisfaction and recommendation quality, the effectiveness of the approach was limited by group size effects, which motivates a follow-up research and the development of a revised method and prototype.
Approach	A new method was developed under the same premise as the previous one: to allow users to collaboratively create and discuss group preferences, from which recommendations are generated and a final one is selected jointly. However, the process by which the group's preference model is created is simplified. Instead of creating individual preference models and negotiating the inclusion/priority of the features in them, users are now able to manipulate the group model directly. The new method consists of two phases: (1) users define and propose individual features, and (2) proposed features must be accepted by the whole group to be included in the group's preference model. Once a feature becomes a group preference, each member may assign an individual importance to it, the aggregation of which defines the feature's overall relevance during the generation of recommendations. Along these lines, a second GRS prototype was implemented to provide hotel recommendations for non-collocated groups (Hootle+) and evaluated in a user study (39 participants) focusing on the effects of group size.
Results	The outcome indicates that the new method improves the scalability of the approach, by which bigger groups showed greater satisfaction and sense of helpfulness when using the system than did the smaller ones. Furthermore, the dichotomy between proposed and accepted features appears to serve as a filter that keeps the attributes in the group model low in number, since the size of group models was consistent through group sizes. On the other hand, a lower success rate per session was obtained for larger groups (i.e. reaching consensus on the final choice), some users were still concerned about the "convoluted" generation of the group model, and there were issues regarding the complexity of the user interface which resulted in a relatively low SUS score.

4.3 Paper III

Table 4.3: Summary of Paper III. Full manuscript on page 113.

	Negotiation and Reconciliation of Preferences in a Group Recommender System
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	Journal of Information Processing – Vol.26, 2018
Type	Journal Article
Status	Published
RQs	RQ1: ● RQ2: ● RQ3: ●
Background	The previously developed method for group decision making supported by a GRS (Paper II) contemplated the inclusion of features proposed by a user into the group model only upon approval by the other group members. However, the additional steps introduced were found complex by some users. Questions arise as to whether the process could be simplified to further improve the transition from individually generated attributes to group preferences, and the importance that private preferences have in the process.
Approach	The negotiation of preferences is redesigned to create a more streamlined process, where group members can add desired features directly into the group model. However, not all the features within the group model are used to generate recommendations: users may give a rating to each attribute, and only those with the top-N highest aggregated ratings are considered. In accordance, a new GRS prototype was implemented to produce hotel recommendations for groups. This time, it was specifically designed for smartphones (Hootle Mobile), which pose a more realistic scenario where group discussion may occur today. A user study was conducted (42 participants) to assess the performance of the more recent method and prototype, also considering different group sizes. Another goal of the evaluation is to obtain a better insight into balancing private and shared areas, as it covers, in conjunction with previous evaluations, a wide range of preference exposure methods: from handling individual preference models only, to their suppression in favour of direct manipulation of group preferences.
Results	In comparison, the more streamlined method achieved higher success rates per session than previous ones, a significantly better SUS score, slightly shorter times per task and group's model size, and was not outperformed by previous methods in any other aspect no matter the size of the group. Therefore, the lack of private preferences does not seem to have any drawbacks in the recommendation process, while it appears to have a positive impact on the user experience. Thanks to the examination of the three elaborated methods, it is possible to more clearly distinguish the different aspects and phases of a group decision process supported by a GRS, which permits the definition of an initial model for such processes.

4.4 Paper IV

Table 4.4: Summary of Paper IV. Full manuscript on page 129.

	Augmented Reality Based Recommending in the Physical World
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	Mensch und Computer 2018 – MuC 2018 Workshop on VR and AR in Everyday Context (VARECo)
Type	Workshop Paper
Status	Published
RQs	RQ4: ◐
Background	Advances in augmented reality technology create new opportunities for applying recommender systems to physical contexts. This possibility, however, has been rarely addressed by current research, despite the benefits that RS may bring in decision-making situations such as physical shopping scenarios, more so when coupled with other decision support functions (e.g. function explanations or comparison tools). Due to the lack of research, it is unclear to what extent AR-based recommendations can support decision-makers in physical settings, and no guidelines exist regarding their successful implementation.
Approach	To obtain initial insight into the capabilities of AR to enable recommending functions, a prototype application was developed able to recognize physical printers and provide content-based recommendations after collecting the customer's preferences, which can be fine-tuned via attribute critiquing. The application was developed for Microsoft's head-mounted display HoloLens, and its implementation allowed experimenting with different interaction (gaze-based selection, air tapping and natural language recognition) and information acquisition (product functionality explained via 3D augmentations, text and text-to-speech) techniques. An embodied virtual advisor was also included to provide guidance through the buying process and give under request information. The prototype was evaluated in a small laboratory user study (15 participants) from which to draw initial design guidelines and usability insights.
Results	Two main conclusion were obtained from the study: first, the number of information sources should be kept as low as possible, fitting on the screen and anchored to real world objects, to avoid breaking the immersion and disorienting the user; and second, the intention to use AR technology seems to depend largely on whether the use of AR is sufficiently justified so as not to use another alternative instead. Under these guidelines, a new concept for AR-based RS is proposed, which removes distracting elements (e.g. embodied virtual advisor) and places a greater focus on using AR advanced visualization capabilities to support decision-making in ways hardly replicable by other means (e.g. aids for physical product comparison).

4.5 Paper V

Table 4.5: Summary of Paper V. Full manuscript on page 137.

	Augmented-Reality-Enhanced Product Comparison in Physical Retailing
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	Mensch und Computer 2019 – MuC 2019
Type	Conference Paper
Status	Published
RQs	RQ4: ◐
Background	The application of AR for utilitarian purposes is still very limited outside of professional contexts. The revolution that physical shopping is undergoing to match the consumer experience to that of digital retailers makes it a promising playground where AR can be applied for practical reasons, as it has the potential of bringing the physical and virtual shopping experience together. In a physical shopping scenario, consumers often rely on physically comparing products and evaluating information obtained from flyers or salespeople. Taking online stores as reference, AR can enable the addition of purchase decision support tools into physical scenarios, such as product comparison aids and extended product information. As this approach has not been previously investigated, it is required to find suitable visualization and interaction methods to support consumers in the exploration of product attributes in a comparative manner.
Approach	A prototype application was designed and developed for Microsoft's HoloLens to learn more about the design requirements of in-store AR-based shopping support tools with comparison capabilities. Taking vacuum cleaners as product domain, the system is able to detect physical products and uses digital augmentations to provide a detailed view of product attributes, including explanations, and to increase the user's awareness of the differences between them. Two user studies (50 and 29 participants) were performed to assess the validity of the approach and to investigate the performance of different attribute comparison visualization methods and interaction techniques.
Results	There was a very positive overall outcome in terms of user experience and satisfaction. Although the inclusion of comparison features had a low impact in that regard, as similar results were obtained by a control group where comparison functions were disabled, quicker information acquisition times were reported when comparison was enabled. When comparing the values of an attribute, their absolute (unmodified) presentation was generally preferred. Implicit attribute selection (through head gaze and a timer) obtained better results than explicit one (via air tapping) in terms of hedonic quality and attribute examination, but requiring a more careful navigation. According to previous studies, explicit activation may be preferred in the long term.

4.6 Paper VI

Table 4.6: Summary of Paper VI. Full manuscript on page 149.

	In-Store Augmented Reality-Enabled Product Comparison and Recommendation
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	14th ACM Conference on Recommender Systems – RecSys 2020
Type	Conference Paper
Status	Published
RQs	RQ4: ◐ RQ5: ◐
Background	Physical retailing lacks the ease with which online stores provide useful data and shopping tools. AR can help bridge this gap by enabling customers to explore product attributes through augmentations and assisting them in the process of understanding their differences. Further extending these functions with recommender technologies has the potential to increase search and decision support, more so if items from the vendor's online offerings are also included in the AR presentation, in line with new multi-channel retailing trends. In this case, physical products can help clients to construct their preferences more effectively and serve as reference points to better understand the attributes of digital ones. Since this concept for in-store AR-based shopping support has not been explored before, the effectiveness of product recommendations provided via AR is still unknown, as are the effects of hybrid shopping context on the way users learn and explore the product space.
Approach	To investigate the aforementioned aspects, an AR-based shopping support system was designed and developed for Microsoft's HoloLens. It combines product comparison and recommending methods for both physical and online products, significantly extending the work described in Paper V. By using the system, users have access to relevant attributes of physical products, and receive recommendations of items similar to the product they are currently inspecting, which can be influenced by critiquing its features. Furthermore, the comparison of physical products against each other and against digital ones is also supported. The prototype was used in a small laboratory study (10 participants) to evaluate the approach and investigate the implications of a hybrid purchase scenario.
Results	The system was positively rated and perceived as useful and intuitive. Physical items influenced how the digital space was browsed, as participants focused on the recommendations given for specific physical products, more specifically those they considered closer to their preferences. Moreover, physical products were regarded as helpful for forming an opinion of the ones available only in digital form. Despite the limitations of the study (the low number of participants, the lab setting and the lack of a baseline), there is enough evidence to consider this to be a viable approach worth to be further explored.

4.7 Paper VII

Table 4.7: Summary of Paper VII. Full manuscript on page 161.

	Acceptance of an AR-Based In-Store Shopping Advisor ***The Impact of Psychological User Characteristics***
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	18th IFIP TC 13 International Conference onHuman-Computer Interaction – INTERACT 2021
Type	Conference Paper
Status	Published
RQs	RQ4: ◐ RQ5: ◐ RQ6: ◐
Background	In the coming era of omni-channel retailing, AR could be used to bring online and physical stores together by enabling the use of digital tools within physical shopping scenarios. This idea was explored in previous research (Paper VI) where an approach for AR-based in-store shopping advisors was presented, capable of providing extended product information, comparison support and product recommendations from physical and digital catalogues. However, it is unclear whether the use of AR technology in such context is acceptable to all users and which psychological characteristics may determine acceptance and attractiveness, particularly when the system involves wearing an AR headset.
Approach	An exploratory study was designed to evaluate the possible implications of psychological characteristics on the acceptance of the approach. Tools to measure relevant traits were combined to create a user profile, more specifically the scales: Technological Adoption Propensity (TAP), Decision Style (DS, intuitive or rational) and Chronic Shopping Orientation (CSO, experiential or task-focused). An online survey was conducted (63 participants) to collect data on personal characteristics (by using the mentioned scales) and user acceptance of the concept, showcased through videos of an improved version of the prototype presented in Paper VI.
Results	There is an indication of the existence of some psychological traits that have an impact on the acceptance of an AR-based in-store advisor. It was possible to determine the presence of four well distinguished types of consumers, who also differ in their acceptance of the system. The results show that technology-related aspects are not the only determinants of AR acceptance, but that other factors are involved as well. The approach is generally well-received, but users with low TAP scores are less less likely to make use of it. Both persons whit high TAP and those with an experiential CSO present higher acceptance values. However, AR knowledge above the average has a moderating effect on the acceptance of users with high TAP values. Most concerns related to privacy and social acceptance factors, the latter being more prominent among users with intuitive DS, although these aspects may become less significant as the technology advances and head-mounted displays become less intrusive.

4.8 Paper VIII

Table 4.8: Summary of Paper VIII. Full manuscript on page 185.

	Creating Omni-Channel In-Store Shopping Experiences through Augmented-Reality-Based Product Recommending and Comparison
Authors	Jesús Omar Álvarez Márquez and Jürgen Ziegler
Venue	International Journal of Human–Computer Interaction – 2023
Type	Journal Article
Status	Published
RQs	RQ4: ● RQ5: ● RQ6: ●
Background	Prior publications introduced a novel approach for enhancing the in-store shopping experience and support the purchase decision by using AR. A scenario is proposed where clients can inspect products physically and digitally by accessing contextual information via superimposed augmentations, which can also be directly compared against that of the other products. Recommendations of similar items are made available too, which are retrieved from the vendor's online and physical catalogues to support the union of both retail channels. Previous evaluations of a prototype system for HMDs indicate benefits in terms of item discovery and decision support, and uncover consumer concerns and target groups. However, these studies are limited by the lack of a baseline, which would allow a better understanding of the real implications of using an AR-based approach for offering in-store services.
Approach	To analyse the real impact of providing in-store AR functions via HMDs, a baseline system with similar functionality but without AR technology was developed for smartphones. An online exploratory study was conducted (64 participants) to continue the work done in Paper VII, with the objective of better identifying consumers types and their view of AR-based and baseline systems. It was followed by a small lab experiment (13 participants) using an updated version of the system in Paper VI to address more specific aspects regarding usability and user experience.
Results	The outcome suggest that the proposed functions create beneficial new dynamics in how consumers learn about, explore, and discover products. The results in Paper VII about the existence of stable consumer types are reinforced, but no significant differences were found between the acceptance of the systems. Providing in-store services via AR HMDs maintains and even improves the pragmatic qualities of using the baseline system. Only the practical aspects of the approach influence the intention to use AR at stores, and ease of use and usability aspects are of great importance for choosing one system or the other. These factors appear to pose a greater challenge in the adoption of AR HMDs than privacy and social acceptance. It is therefore more urgent to focus on defining practical uses for AR and suitable visualization and interaction methods to make HMDs competitive against more standard displays.

5 Conclusions and Future Work

5.1 Results

Table 5.1 shows a summary of research questions addressed by each published paper, as well as the contributions of each one of them.

Table 5.1: Research questions and contributions per paper.

	Papers							
	I	II	III	IV	V	VI	VII	VIII
Complex multi-actor decision environments								
RQ1 Information sharing and negotiation in GRS	◐	◐	●					
RQ2 Private vs shared information spaces	◐	◐	●					
RQ3 Usability and acceptance of negotiation-based GRS	◐	◐	●					
Contributions:								
Preference elicitation and aggregation methods	•	•	•					
Information sharing and visualization methods	•	•	•					
Effects of group size on consensus building	•	•	•					
Negotiation-driven GRS prototype implementation	•	•	•					
Model for group decision supported by a GRS			•					
Complex multi-source decision environments								
RQ4 Advantages of providing in-store AR-based functions				◐	◐	◐	◐	●
RQ5 Implications of a hybrid physical-digital setting						◐	◐	●
RQ6 Effects of consumer psychological traits							◐	●
Contributions:								
In-store AR-based product comparison methods				•	•	•		
In-store AR-based product recommending methods				•		•		
AR-based shopping support system implementation				•	•	•		
Effects on the purchase decision					•	•	•	•
Product exploration/discovery/learning in hybrid settings						•	•	•
Consumer types and their acceptance of the system							•	•
Comparison against a non-AR baseline prototype								•
Evaluations								
Laboratory study	•	•	•	•	•	•		•
Online study							•	•

◐ RQ partially addressed ● RQ completely addressed • Outcome adds to the contribution

Complex multi-factor decision environments can occur in a wide range of situations, and may comprise a variety of factor combinations. For the purposes of this thesis, we have focused on two specific scenarios, which we believe are good representatives of the challenges that modern-day decision-makers may face, and in which the use of technology is a source of complexity. Because of the unbridgeable differences between multi-actor and multi-source decision environments, specific research questions were elaborated and investigated for each of scenario. Therefore, the investigation was divided in two independent lines of research, the results of which shed light on the particular processes that take place in each setting, and define viable support methods appropriate to them. In the following, each research question will be analysed separately to discuss the related results, and reference will be made to the pertinent papers where necessary.

5.1.1 Complex multi-actor decision environments

The research of complex multi-actor decision environments was designed in an iterative manner. The first cycle (Paper I) comprised an initial conception stage for the definition of methods to support the visualization and sharing of information, the elicitation of individual and group preferences, and the generation of recommendations, all of them from a group negotiation and consensus building point of view. The concept was implemented in a prototype, which was evaluated in a user study in order to answer the formulated research questions. The subsequent iterations (Papers II, III) were constructed on the results of previous ones, always focusing on the same core elements but applying modifications based on the observed issues. The results that follow summarize the outcomes obtained after three cycles, for which three different systems were developed and evaluated, each applying its own negotiation, visualization and preference aggregation methods.

RQ1: What are effective means for supporting the formulation, the exchange and the negotiation of group preferences in multi-actor decision-making?

An initial model to support complex multi-actor decision environments via GRS can be draw thanks to the results obtained after the evaluation of different approaches for the negotiation and reconciliation of group preferences (Paper III). The model outlines the phases that group members must go through to reach consensus, where each stage involves cognitive aspects that can be supported by the system.

1. **Development of individual preferences** Users make themselves aware of their preferences, express them, reflect on them and potentially adapt them either

based on their own insight or through interaction with other group members. To this end, the system must provide means for users to explore and learn about product and attribute spaces. The system should also be transparent about the current set of individual preferences, and allow users their direct inspection and modification.

2. **Exchange of preferences** Users reveal and communicate their preferences to other group members or the whole group, either as complete preference profiles or as single feature preferences. The system should offer tools to easily navigate through shared preferences, and include indications about their significance for individual members, and their impact on the group model.

3. **Negotiation of preferences** Group members discuss, criticize, or weight the individual preferences or the group model as a whole, possibly involving voting mechanisms to decide on the acceptability of individual preferences. Functions to ease assessing, discussing, and accepting specific preferences should be included in the system, such as links to make quick chat references, discussion threads, or other methods to share the opinion given by a user to the preferences defined by others.

4. **Evaluation and selection of alternatives** Group members weight, criticize or vote the resulting recommendations, converging on a joint decision. At this stage, the system should provide an updated set of recommendations based on the current state of the group model. Users must be able to explore the given alternatives and propose interesting ones to the group. Functions are required to assess the acceptance of proposed alternatives, as well as to allow their discussion.

It is also suggested to use a two-step process during the generation of recommendations in order to provide a diverse set of alternatives, reduce individual discontent, and foster group discussion and negotiation. After retrieving an initial set of alternatives based on the group model, a second "feature balancing" step can take place to adjust the final set of recommendations to contain at least one fitting item per attribute in the preference model. In this manner, group members with less popular preferences are also given the opportunity to propose alternatives that are interesting for them.

RQ2: How to structure the preference elicitation and construction process into private spaces for setting up one's individual preferences versus public spaces that can be seen and criticized by the whole group?

Three different approaches to group preference elicitation have been explored in this research (Figure 5.1). In all of them, users have a private space where they can explore the available features and configure them to their liking. However, the approaches differ in how the selected attributes become part of the group's preference model.

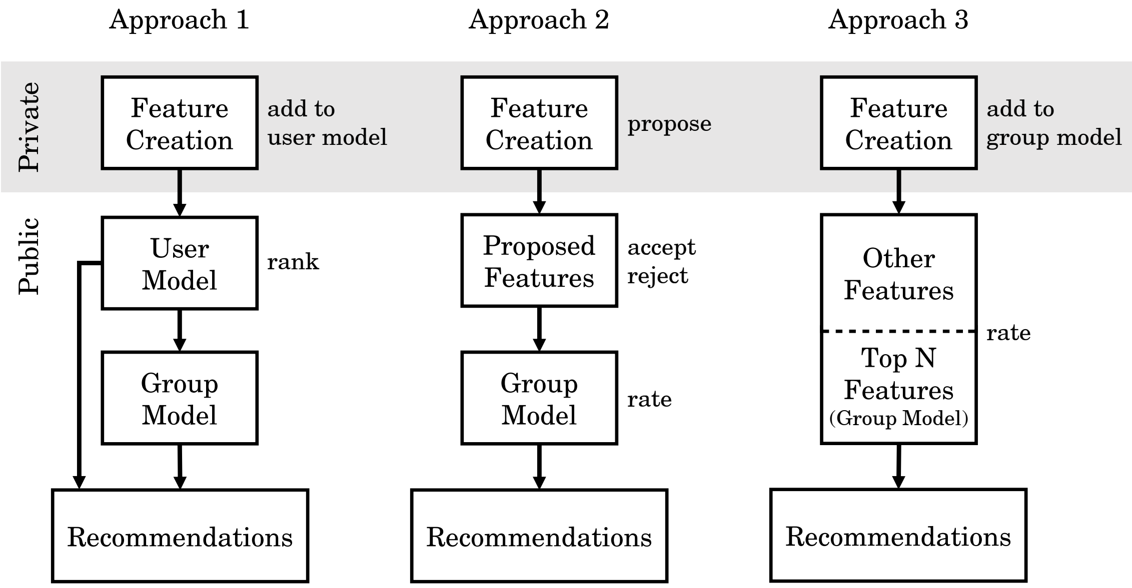


Figure 5.1: Approaches to group preference elicitation. Each box represents a stage in the process. Text next to a box indicates the main actions that can be performed by users.

Approach 1: Negotiating individual preference profiles (Paper I) Attributes selected by a member are added into the user preference model. This model is publicly viewable, but can only be modified by its owner. An aggregation function is used to create and update the group model, which can only be influenced by asking other users to change their individual ones.

Approach 2: Negotiating group preferences (Paper II) Attributes selected by a member are added into a public pool. Other members must vote to either accept or reject a certain attribute into the group model. The accepted attributes become part of the group's preference model, where group members can assign an individual relevance score to each attribute, which is then aggregated into its overall importance within the model. To modify the group model, negotiation must occur, first, during the acceptance of preferences, and second, during the assignment of individual relevance.

Approach 3: Negotiating the relevance of preferences (Paper III) Attributes selected by a member are directly added into the group's preference model. Members can assign an individual relevance score to each attribute, which is then aggregated into its overall importance within the model. Attributes are ranked accordingly, and only the top N ones are used for generating recommendations. Members are aware of it, and must negotiate individual relevance scores to influence the group model.

Each approach achieves a different balance between private and public spaces, which affects what users can do, and what they must ask others to do, during each phase of the group's preference elicitation process. As a consequence, also the number of interaction steps required to complete the process is different between approaches.

The outcome of the research shows that the method that requires the fewer steps (that is, directly negotiating the relevance of preferences) is also the most efficient, and results in a higher perception of the quality of the recommendations. In addition, this is also method with the highest group scalability, because it offers a more centralized view of all preferences, in contrast to the other methods in which keeping track of information is more difficult, either because individual preference models are displayed separately (approach 1), or because public preferences are scattered over different sections (proposed versus accepted attributes in approach 2).

RQ3: How do the proposed approach and developed system perform in terms of usability and acceptance?

In contrast to a non discussion-enabled GRS, the inclusion of discussion and negotiation features has a positive effect on the overall user satisfaction and perceived quality of the recommendations, and offers increased probabilities for reaching group consensus (Paper I). In addition, the assessment of both user satisfaction and recommendation quality became more positive with each new iteration. That is, of the three implemented systems, the more recent one, directly negotiating the relevance of preferences, received the best results overall (Paper III), reaching a usability rate of "excellent" (Bangor et al., 2009). Larger groups had more trouble finding consensus due to an increased complexity of the decision-making process, but appeared to be more satisfied with the provided tools than smaller ones, and made a more extensive use of the graphical interface for expressing their preferences (instead of using the chat).

5.1.2 Complex multi-source decision environments

The research of complex multi-source decision environments was designed in an incremental manner. After the conception of functions to offer in-store shopping support via AR (Paper IV), an initial system with limited functionality was developed for head-mounted displays (HMD), and evaluated in a user study (Paper V). That system was redesigned and further extended and evaluated in following research (Papers VI, VII). While initial papers are more concerned with the interaction, visualization and navigation aspects of the approach (Papers V, VI), the latter are more focused on the investigation of the acceptance of such types of systems by considering, first, the effects of consumer psychological traits (Papers VII, VIII), and second, the system's performance against a non-AR baseline approach for smartphones (Paper VIII). The results that follow summarize the outcomes obtained after multiple laboratory and online evaluations.

RQ4: How can AR-based functions in a physical shopping scenario support the decision-making process in relation to information access, product discovery and purchase confidence?

Study participants made extensive use of AR-enabled comparison and recommending features, which gives a first indication that the tools were perceived as useful during the decision-making process (Paper VI). On the one hand, allowing access to product information and comparison in physical settings appears to allow for a quicker acquisition of information (Paper V), even in juxtaposition to a more standard non-AR smartphone system (Paper VIII); on the other hand, providing in-store recommendations, critiquing possibilities, and access to a digital catalogue seems beneficial for discovering new and diverse products (Papers VI, VII, VIII). Users highly regarded access to joint digital and physical information through AR to make the final purchase decision, get a better view of the product space, and perform physical comparison of products, especially when a large number of alternatives are available (Papers VII, VIII). In this respect, not losing focus from the physical space that surrounds the user seems to encourage product inspection and a more involved shopping process (Paper VIII). Finally, an AR-based approach also brings added entertainment value during the shopping experience (Paper VIII).

As possible limitations, the use of HMDs is generally perceived as a more event-oriented approach (e.g. trade fairs or marketing actions), and issues may arise regarding social acceptance. In a physical store situation, a greater appreciation of the practical value of using AR seems critical for its preference over more standard technological alternatives (Paper VIII). Additionally, it is still required to find adequate methods to better explain the recommendations and explore the attributes in AR environments, in order to get the most out of the approach (Paper VII).

RQ5: What are the implications that a hybrid physical-digital setting may have on the way users learn and explore the digital space?

The evaluation of the prototype systems suggest that a hybrid physical-digital environment (i.e. where products from a digital catalogue can be compared to, and explored through, the physical ones) has effects on aspects related to the information search and evaluation stages of the purchasing process:

Physical products serve as filters for the digital space (Paper VI) Consumers in a hybrid setting make use of the physically present products to explore digital recommendations. Since these recommendations are similar to the product they are attached to, users tend to first find a physical product that is close to their needs, and then focus on exploring the linked recommendations. In this way, clients seem to intuitively reduce the number of available alternatives to only those that already meet certain criteria, and then narrow them down even further via attribute critiquing.

Physical products showcase the digital ones (Papers VI, VII, VIII) When inspecting digital products, participants find useful to utilize physically available ones as a reference to better understand their attributes, and thus obtain a more realistic perception of the qualities of digital-only alternatives. This facilitates forming purchase preferences and the evaluation of suitable options, which ultimately leads to a greater purchase confidence.

As for the use of AR to enable shopping support functions in hybrid environments, the technology appears to provide a better view of the product space and physical comparison, and users are more motivated to physically explore products than when presenting digital information on a separated, standard display (Paper VIII)

RQ6: What is the significance of consumer psychological factors in the adoption of AR-based shopping support systems?

The collection of psychological data from a total of 116 participants in two online user studies made it possible to define four types of consumers according to their propensity to adopt technology, their shopping behaviour and the way they make decisions (Table 5.2).

Table 5.2: Overview of consumer types and their level of acceptance of in-store AR-based functions (relative to the other types).

Type	Traits	Acceptance
1	Experiential shoppers, intuitive	High
2	Technology adopters, highest knowledge about AR	Medium
3	Technology rejectors, lowest knowledge about AR, least rational	Low
4	Task-focused shoppers, most rational, least intuitive	High

The analysis of their assessment of the approach indicates that there are differences in their acceptance of AR-based shopping support systems (Papers VII, VIII):

- Experiential shoppers (Type 1) present a higher acceptance of AR-based in-store shopping support functions, which may be directly related with the hedonic value of AR technology and its effects on the overall shopping experience (Poushneh & Vasquez-Parraga, 2017). However, they also show more pronounced concerns about privacy and social acceptance.
- Results show that higher technology adoption propensity generally correlates with a higher acceptance of the proposed functions. Nonetheless, it also appears that the more participants know about AR (as it is the case for Type 2 consumers), the lower they rate the AR-based approach; that is, AR knowledge has a moderating effect on the acceptance of AR-based functions. This may happen because most users are only exposed to AR for entertainment reasons, which may increase their disbelief in the suitability of the technology for more practical purposes.
- Consumers who do not trust technology nor are interested in new technological advances (Type 3) also present lower acceptance of in-store support. Their reluctance to use in-store functions is generalized, as they score AR and non-AR systems similarly low.
- Task-focused shoppers with a marked rational decision-making style (Type 4) express a more favourable perception of the utilitarian value of AR to access information, compare products and discover new ones, and therefore show a higher acceptance of the proposed AR-based in-store functions.

All consumer types agree on the better suitability of AR for special events like marketing actions, rather than daily shopping activities. There is as well a widespread feeling of concern for social acceptance, mostly due to the conspicuousness of current HMDs.

5.2 Discussion

The study of specific complex multi-factor decision environments allowed to determine their particular characteristics and challenges, as well as opened the possibility to develop and test solutions to support decision-makers in these situations.

Non-collocated multi-actor decision environments must cope with difficulties in the transmission of information, balance of preferences and selection of alternatives. Data shows that the biggest challenge comes from the capability of group members to express their own preferences and evaluate those of others. Balancing the importance of individual and group preferences is a complex matter, for which many approaches can be conceived depending on how the group's preference model is created and how much agency members have to modify it. Supporting the group decision-making process with a GRS and discussion and negotiation techniques seems to alleviate the aforementioned issues. The study of different methods for the negotiation of preferences and consensus building allowed the elaboration of an initial model for this type of systems. From the methods proposed in this research, the more streamlined one together with a simplified user interface appears to provide the best results on the user experience. The final, simplified process resulting of the investigation received good scores for usability-related criteria, scalability for larger group sizes, and perceived recommendation quality in the case of larger groups. The proposed model for supporting multi-actor decision environments through GRS, along with the design lessons learned from the iterative development process, provide a solid foundation on which to build new research.

With respect to a multi-source decision environment, people in such situations must face the challenge of compiling, filtering and evaluating alternatives from different information sources, which may differ in aspects such as accessibility, availability or reliability of the provided data. When dealing with hybrid environments where digital and physical information coexist, the extent to which the decision-maker is able to integrate and navigate through both information spheres has effects on several stages of the decision-making process. Within the more specific case of making a purchase decision, the obtained results indicate that consumers benefit from having access to comparison and recommendation functions that support the acquisition, understanding and evaluation of information in physical settings. Users also seem to intuitively adapt to the specific characteristics of a hybrid environment supported by these functions, and tend to use physical elements to explore and learn about digital ones. Furthermore, providing such functions through AR HMDs yields greater value in terms of product exploration and discovery than a non-AR smartphone approach, while there is no significant loss regarding other pragmatic factors but exists a clear advantage in hedonic ones. Nonetheless, using an AR HMD poses a higher level of complexity, and further technological advances and research on visualization and interaction techniques may be necessary to increase ease of use and decrease social acceptance concerns.

The developed approaches demonstrate the capability of current technology to support decision-making in multi-actor and multi-source situations. Although adapted to each scenario, the use of recommendations resulted beneficial in either case. The value of recommender systems in the decision-making process is evident, as is their great capacity to adapt to each circumstance.

Limitations Although the results of studying the proposed scenarios can be partially extrapolated to more general multi-actor or multi-source decision environments, the developed methods have been designed to support decision-makers in very specific situations. Both multi-actor and multi-source decision environments can take a wide range of forms, and it is likely for each one to have its own particularities and require unique solutions. In addition, this work has focused on two specific configurations of complex multi-factor decision environments, but there is a third possibility that has not been contemplated here, which results from the combination of multiple actors and sources.

5.3 Conclusions and Outlook

In order to gain a better understanding of the main processes occurring in complex multi-factor decision environments and how to support them, this dissertation has focused on the independent investigation of two specific multi-actor and multi-source scenarios. Insight into their characteristics and limitations was gained by studying previous literature and conducting empirical evaluations. Support methods were designed for each scenario that make use of recommender systems and new technological advances. Their implementation into functional prototypes and subsequent evaluation in multiple user studies has allowed a better comprehension of the challenges that decision-makers must face, and the implications of the proposed solutions in that regard.

Non-collocated complex multi-actor decision environments find benefits in using GRS that support discussion and consensus building. A model describing the phases that a group goes through during the negotiation and reconciliation of preferences is proposed in this dissertation, and design guidelines for negotiation-based GRS can be extracted from the evaluation of the implemented prototypes.

Complex multi-source purchase decision environments in physical settings can take advantage of product recommendations afforded by the use of AR-technology. A concept for providing unified information access, product comparison and recommending functions via augmentations was investigated, the evaluation of which provides valuable data about the suitability of the concept, its performance in contrast to more standard information acquisition methods, and its acceptance among consumer types.

The results of this research open several paths for further investigation. First, it is possible to continue the work done in each one the proposed scenarios. The approach here described for complex multi-actor decision-making environments remains unique within its area of research (Alvarado Rodriguez et al., 2022), which offers a great opportunity to keep exploring it. Similarly, decision support in hybrid physical-digital environments is a novel concept that relies heavily on new technological developments and requires constant adaptation to the latest innovations. Second, the possibility exists to explore other complex multi-actor and multi-source decision environments that occur in completely different scenarios, and for which finding specific solutions may be required. This could allow for a more general abstraction of processes and support methodologies for multi-factor settings. Finally, the combination of multiple actors and sources has been purposely overlooked by this research, but it could be interesting to further explore its implications.

Bibliography

Abbar, S., Bouzeghoub, M. & Lopez, S. (2009). *Context-aware recommender systems: A service-oriented approach*. In *VLDB PersDB workshop*, pp. 1–6. 26

Adomavicius, G. & Tuzhilin, A. (2005). *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions*. IEEE transactions on knowledge and data engineering, 17(6):734–749. 19

Ahn, J., Williamson, J., Gartrell, M., Han, R., Lv, Q. & Mishra, S. (2015). *Supporting Healthy Grocery Shopping via Mobile Augmented Reality*. ACM Trans. Multimedia Comput. Commun. Appl., 12(1s):16:1–16:24. 18

Aisle 411 & Tango (2014). *InStore 3D mapping at Walgreens*. Online: http://aisle411.com/walgreens/; accessed 12-05-2022. 27

Alvarado Rodriguez, O. L., Htun, N. N., Jin, Y. & Verbert, K. (2022). *A systematic review of interaction design strategies for group recommendation systems*. Proceedings of the ACM on Human-Computer Interaction. 3, 24, 25, 31, 56

Amer-Yahia, S., Elbassuoni, S., Omidvar-Tehrani, B., Borromeo, R. & Farokhnejad, M. (2019). *Grouptravel: Customizing travel packages for groups*. In *International Conference on Extending Database Technology (EDBT) 2019*. 22

Ardissono, L., Goy, A., Petrone, G., Segnan, M. & Torasso, P. (2003). *Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices*. Applied artificial intelligence, 17(8-9):687–714. 22, 23

Arnott, D. & Pervan, G. (2015). *A critical analysis of decision support systems research*. In *Formulating Research Methods for Information Systems: Volume 2*, pp. 127–168. Palgrave Macmillan UK, London. 1, 3, 6

Azuma, R. T. (1997). *A Survey of Augmented Reality*. Presence: Teleoperators and Virtual Environments, 6(4):355–385. 18

Baker, D., Bridges, D., Hunter, R., Johnson, G., Krupa, J., Murphy, J. & Sorenson, K. (2001). *Guidebook to decision-making methods*, vol. 45. Westinghouse Savannah River Company Aiken, SC, USA. 12, 24

Baltes, B. B., Dickson, M. W., Sherman, M. P., Bauer, C. C. & LaGanke, J. S. (2002). *Computer-mediated communication and group decision making: A meta-analysis*. Organizational behavior and human decision processes, 87(1):156–179. 8

Bangor, A., Kortum, P. & Miller, J. (2009). *Determining what individual SUS scores mean: Adding an adjective rating scale*. Journal of usability studies, 4(3):114–123. 51

Basu Roy, S., Amer-Yahia, S., Chawla, A., Das, G. & Yu, C. (2010). *Space efficiency in group recommendation*. The VLDB Journal, 19(6):877–900. 24

Baxendale, S., Macdonald, E. K. & Wilson, H. N. (2015). *The impact of different touchpoints on brand consideration*. Journal of Retailing, 91(2):235–253. 13

Beck, N. & Rygl, D. (2015). *Categorization of multiple channel retailing in Multi-, Cross-, and Omni-Channel Retailing for retailers and retailing*. Journal of Retailing and Consumer Services, 27:170–178. 14, 18, 25, 34

Beckmann, C. & Gross, T. (2010). *Towards a group recommender process model for ad-hoc groups and on-demand recommendations*. In *Proceedings of the 16th ACM international conference on Supporting group work*, pp. 329–330. 23, 24

Bennet, A. & Bennet, D. (2008). *The decision-making process in a complex situation*. In *Handbook on Decision Support Systems 1*, pp. 3–20. Springer. 7

Berry, G. R. (2006). *Can computer-mediated asynchronous communication improve team processes and decision making? Learning from the management literature*. The Journal of Business Communication (1973), 43(4):344–366. 1, 12, 13, 31

Berry, G. R. (2011). *Enhancing effectiveness on virtual teams: Understanding why traditional team skills are insufficient*. The Journal of Business Communication (1973), 48(2):186–206. 8, 13, 31

Blackwell, R. D., Miniard, P. W. & Engel, J. F. (2006). *Consumer behavior*. South-Western Pub. 14, 17

Bobadilla, J., Ortega, F., Hernando, A. & Gutiérrez, A. (2013). *Recommender systems survey*. Knowledge-based systems, 46:109–132. 19, 22, 23

Bonetti, F., Pantano, E., Warnaby, G., Quinn, L. & Perry, P. (2019). *Augmented reality in real stores: empirical evidence from consumers' interaction with AR in a retail format*. In *Augmented Reality and Virtual Reality*, pp. 3–16. Springer. 18

Borda, J. d. (1784). *Mémoire sur les élections au scrutin*. Histoire de l'Academie Royale des Sciences pour 1781 (Paris, 1784). 11

Bressen, T. (2007). *Consensus decision making*. The change handbook: The definitive resource on today's best methods for engaging whole systems, 495:212–217. 11, 30, 31

Broeckelmann, P. & Groeppel-Klein, A. (2008). *Usage of mobile price comparison sites at the point of sale and its influence on consumers' shopping behaviour*. The International Review of Retail, Distribution and Consumer Research, 18(2):149–166. 2

Bubnicki, Z. (2013). *Analysis and decision making in uncertain systems*. Springer Science & Business Media. 8

Burke, R. (2000). *Knowledge-based recommender systems*. Encyclopedia of library and information systems, 69(Supplement 32):175–186. 19

Candillier, L., Meyer, F. & Boullé, M. (2007). *Comparing state-of-the-art collaborative filtering systems*. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 548–562. Springer. 19

Cappel, J. J. & Windsor, J. C. (2000). *Ethical decision making: A comparison of computer-supported and face-to-face group*. Journal of Business Ethics, 28(2):95–107. 12, 13, 31

Chao, D. L., Balthrop, J. & Forrest, S. (2005). *Adaptive radio: achieving consensus using negative preferences*. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pp. 120–123. 22

Chen, C.-C., Huang, T.-C., Park, J. J. & Yen, N. Y. (2015). *Real-time smartphone sensing and recommendations towards context-awareness shopping*. Multimedia Systems, 21(1):61–72. 26

Chen, Y.-C., Shang, R.-A. & Kao, C.-Y. (2009). *The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment*. Electronic Commerce Research and Applications, 8(1):48–58. 2, 8, 19, 25, 33, 34

Church, K., Smyth, B., Cotter, P. & Bradley, K. (2007). *Mobile information access: A study of emerging search behavior on the mobile internet*. ACM Transactions on the Web (TWEB), 1(1):4–es. 2, 17

Chylinski, M., De Ruyter, K., Sinha, A. & Northey, G. (2014). *Augmented retail reality: Situated cognition for healthy food choices*. ANZMAC (Australian and New Zealand Marketing Academy). 18

Citroen, C. L. (2011). *The role of information in strategic decision-making*. International journal of information management, 31(6):493–501. 2, 6, 8

Clark, D. (2013). *Using social media to map the consumer journey to the customer experience*. MyCustomer. 16

Copeland, A. H. (1951). *A reasonable social welfare function*. Technical Report, mimeo, 1951. University of Michigan. 11

Court, D., Elzinga, D., Mulde, S. & Vetvik, O. J. (2009). *The consumer decision journey*. McKinsey Quarterly. 16, 18

Cox, A., Granbois, D. & Summers, J. (1983). *Planning, search, certainty and satisfaction among durables buyers: a longitudinal study*. ACR North American Advances. 14

Crossen, A., Budzik, J. & Hammond, K. J. (2002). *Flytrap: intelligent group music recommendation*. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pp. 184–185. 22

Cruz, E., Orts-Escolano, S., Gomez-Donoso, F., Rizo, C., Rangel, J. C., Mora, H. & Cazorla, M. (2019). *An augmented reality application for improving shopping experience in large retail stores*. Virtual Reality, 23(3):281–291. 3, 18

Dacko, S. G. (2017). *Enabling smart retail settings via mobile augmented reality shopping apps*. Technological Forecasting and Social Change, 124:243–256. 18

Damian, D. E. & Zowghi, D. (2002). *The impact of stakeholders' geographical distribution on managing requirements in a multi-site organization*. In *Proceedings IEEE Joint International Conference on Requirements Engineering*, pp. 319–328. IEEE. 2, 8, 31

Darley, W. K., Blankson, C. & Luethge, D. J. (2010). *Toward an integrated framework for online consumer behavior and decision making process: A review*. Psychology & marketing, 27(2):94–116. 15

Edwards, W. & Fasolo, B. (2001). *Decision technology*. Annual review of psychology, 52(1):581–606. 5

Eom, S. (1997). *Assessing the current state of intellectual relationships between the decision support systems area and academic disciplines*. ICIS 1997 Proceedings, p. 11. 3, 6

Fagerstrøm, A., Eriksson, N. & Sigurdsson, V. (2020). *Investigating the impact of Internet of Things services from a smartphone app on grocery shopping*. Journal of Retailing and Consumer Services, 52:101927. 26

Far, B. H. & Wahono, R. S. (2003). *Cognitive-decision-making issues for software agents*. Brain and Mind, 4(2):239–252. 5

Felfernig, A., Boratto, L., Stettinger, M. & Tkalčič, M. (2018). *Group recommender systems: An introduction*. Springer. 2, 22, 31

Gauvain, M. & Rogoff, B. (1989). *Collaborative problem solving and children's planning skills.*. Developmental psychology, 25(1):139. 9

Gokhale, A. A. (1995). *Collaborative learning enhances critical thinking*. 9

Gutiérrez, F., Htun, N. N., Charleer, S., De Croon, R. & Verbert, K. (2019). *Designing Augmented Reality Applications for Personal Health Decision-Making*. 3, 18

Haas, M. R., Criscuolo, P. & George, G. (2015). *Which Problems to Solve? Online Knowledge Sharing and Attention Allocation in Organizations*. Academy of Management Journal, 58(3):680–711. 2, 6, 8

Halvorsrud, R., Kvale, K. & Følstad, A. (2016). *Improving service quality through customer journey analysis*. Journal of service theory and practice. 13

Harris, R. (1998). *Introduction to Decision Making*. VirtualSalt. 5, 16

Harrison, E. F. (1999). *The managerial decision-making process*. Houghton Mifflin College Division. 5

Hermes, A., Sindermann, C., Montag, C. & Riedl, R. (2022). *Exploring Online and In-Store Purchase Willingness: Associations With the Big Five Personality Traits, Trust, and Need for Touch*. Frontiers in Psychology, 13. 27

Hilbert, M. & López, P. (2011). *The World's Technological Capacity to Store, Communicate, and Compute Information*. Science. 2, 8

Hilken, T., Heller, J., Chylinski, M., Keeling, D. I., Mahr, D. & de Ruyter, K. (2018). *Making omnichannel an augmented reality: the current and future state of the art*. Journal of Research in Interactive Marketing, 12(4):509–523. 14, 18

Hinds, P., Kiesler, S. B. & Kiesler, S. (2002). *Distributed work*. MIT press. 12

Hirokawa, R. Y. (1988). *Group communication and decision-making performance: A continued test of the functional perspective*. Human Communication Research, 14(4):487–515. 10

Hogarth, R. M. & Makridakis, S. (1981). *The value of decision making in a complex environment: An experimental approach*. Management Science, 27(1):93–107. 7

Howard, J. A. & Sheth, J. N. (1969). *The theory of buyer behavior*. New York, 63:145. 14

Hoyer, W. D. (1984). *An examination of consumer decision making for a common repeat purchase product*. Journal of consumer research, 11(3):822–829. 18

Huré, E., Picot-Coupey, K. & Ackermann, C.-L. (2017). *Understanding omni-channel shopping value: A mixed-method study*. Journal of Retailing and Consumer Services, 39:314–330. 14

Huynh, B., Ibrahim, A., Chang, Y. S., Hollerer, T. & O'Donovan, J. (2018). *A Study of Situated Product Recommendations in Augmented Reality*. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE. 27

Jameson, A. (2004). *More than the sum of its members: challenges for group recommender systems*. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 48–54. 23, 24

Javornik, A., Rogers, Y., Moutinho, A. M. & Freeman, R. (2016). *Revealing the Shopper Experience of Using a "Magic Mirror" Augmented Reality Make-Up Application*. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS 16*. ACM Press. 27

Johnson, D. W. & Johnson, F. P. (1991). *Joining together: Group theory and group skills*. Prentice-Hall, Inc. 10

Jonassen, D. H. & Kwon, H. (2001). *Communication patterns in computer mediated versus face-to-face group problem solving*. Educational technology research and development, 49(1):35–51. 12, 13

Jugovac, M. & Jannach, D. (2017). *Interacting with recommenders—overview and research directions*. ACM Transactions on Interactive Intelligent Systems (TiiS), 7(3):1–46. 21

Kang, J.-Y. M. (2018). *Showrooming, webrooming, and user-generated content creation in the omnichannel era*. Journal of Internet Commerce, 17(2):145–169. 14, 34

Kelly, D. & Teevan, J. (2003). *Implicit feedback for inferring user preference: a bibliography*. In *Acm Sigir Forum*, vol. 37, pp. 18–28. ACM New York, NY, USA. 21

Khusro, S., Ali, Z. & Ullah, I. (2016). *Recommender systems: issues, challenges, and research opportunities*. In *Information science and applications (ICISA) 2016*, pp. 1179–1189. Springer. 21

Kim, J. & Forsythe, S. (2008). *Adoption of Virtual Try-on technology for online apparel shopping*. Journal of Interactive Marketing, 22(2):45–59. 27

Kim, J. S. & Ratchford, B. T. (2012). *Consumer choice and use of multiple information sources for automobile purchases*. International Journal of Electronic Commerce, 16(3):7–40. 17

Klein, L. R. & Ford, G. T. (2003). *Consumer search for information in the digital age: An empirical study of prepurchase search for automobiles*. Journal of interactive Marketing, 17(3):29–49. 17, 33

Kleinmuntz, D. N. (1985). *Cognitive heuristics and feedback in a dynamic decision environment*. Management science, 31(6):680–702. 7

Kocas, C. (2002). *Evolution of Prices in Electronic Markets Under Diffusion of Price-Comparison Shopping*. Journal of Management Information Systems, 19(3):99–119. 25

Konuş, U., Verhoef, P. C. & Neslin, S. A. (2008). *Multichannel shopper segments and their covariates*. Journal of retailing, 84(4):398–413. 15

Kotler, P., Keller, K. L., Ancarani, F. & Costabile, M. (2014). *Marketing management 14/e*. Pearson. 14

Kourouthanassis, P., Spinellis, D., Roussos, G. & Giaglis, G. M. (2002). *Intelligent cokes and diapers: MyGrocer ubiquitous computing environment*. In *First International Mobile Business Conference*, pp. 150–172. 25, 26

Kunaver, M. & Požrl, T. (2017). *Diversity in recommender systems–A survey*. Knowledge-based systems, 123:154–162. 19, 21, 22

Lackermair, G., Kailer, D. & Kanmaz, K. (2013). *Importance of Online Product Reviews from a Consumer's Perspective*. 25

Lavidge, R. J. & Steiner, G. A. (1961). *A model for predictive measurements of advertising effectiveness*. Journal of marketing, 25(6):59–62. 14

Lee, Y., Kozar, K. A. & Larsen, K. R. (2003). *The technology acceptance model: Past, present, and future*. Communications of the Association for information systems, 12(1):50. 27

Levy, D. M. (2008). *Information overload*. The handbook of information and computer ethics, p. 497. 8

Lieberman, H., Van Dyke, N. W. & Vivacqua, A. S. (1998). *Let's browse: a collaborative Web browsing agent*. In *Proceedings of the 4th international conference on Intelligent user interfaces*, pp. 65–68. 23

Lu, A. C. C. & Gursoy, D. (2015). *A conceptual model of consumers' online tourism confusion*. International Journal of Contemporary Hospitality Management. 2, 18

Ludwig, T., Hoffmann, S., Jasche, F. & Ruhrmann, M. (2020). *VacuumCleanAR: augmented reality-based self-explanatory physical artifacts*. In *Proceedings of the Conference on Mensch und Computer*, pp. 291–302. 3, 18

Lunenburg, F. C. (2010a). *The decision making process.*. In *National Forum of Educational Administration & Supervision Journal*, vol. 27. 5

Lunenburg, F. C. (2010b). *Group decision making*. In *National Forum of Teacher Education Journal*, vol. 20, pp. 1–7. 9, 24, 30, 31

Lye, A., Shao, W., Rundle-Thiele, S. & Fausnaugh, C. (2005). *Decision waves: consumer decisions in today's complex world*. European Journal of Marketing. 18

Maity, M., Dass, M. & Malhotra, N. K. (2014). *The antecedents and moderators of offline information search: A meta-analysis*. Journal of Retailing, 90(2):233–254. 17

Marlow, S. L., Lacerenza, C. N. & Salas, E. (2017). *Communication in virtual teams: A conceptual framework and research agenda*. Human Resource Management Review, 27(4):575–589. 2, 8, 13, 31

Masthoff, J. (2004). *Group modeling: Selecting a sequence of television items to suit a group of viewers*. In *Personalized digital television*, pp. 93–141. Springer. 10, 22, 23

McCarthy, J. F. & Anagnost, T. D. (1998). *MusicFX: an arbiter of group preferences for computer supported collaborative workouts*. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pp. 363–372. 22, 23

McCarthy, K., Salamó, M., Coyle, L., McGinty, L., Smyth, B. & Nixon, P. (2006). *Cats: A synchronous approach to collaborative group recommendation*. In *Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pp. 86–91. 23, 24

Meier, J. V., Noel, J. A. & Kaspar, K. (2021). *Alone together: computer-mediated communication in leisure time during and after the COVID-19 pandemic*. Frontiers in Psychology, 12:2040. 1, 12

Mix, L. & Katzberg, J. (2015). *Retails next top model: new perspectives on consumer decision processes in a rapidly changing world*. ESOMAR Qualitative, Paris, November, 18:2018. 2, 18

Molenaar, C. (2016). *Shopping 3.0: Shopping, the internet or both?*. Routledge. 15

Perea y Monsuwé, T., Dellaert, B. G. & de Ruyter, K. (2004). *What drives consumers to shop online? A literature review*. International Journal of Service Industry Management, 15(1):102–121. 25

Mora, D., Jain, S., Nalbach, O. & Werth, D. (2020). *Holographic Recommendations in Brick-and-Mortar Stores*. 27

Moustakas, C. (1990). *Heuristic research: Design, methodology, and applications*. Sage Publications. 6

Nguyen, T. N. & Ricci, F. (2018). *A chat-based group recommender system for tourism*. Information Technology & Tourism, 18(1):5–28. 24

Nielsen, H. (2011). *Bounded rationality in decision making*. Dobbs Ferry, NY. 1, 6

O'connor, M., Cosley, D., Konstan, J. A. & Riedl, J. (2001). *PolyLens: A recommender system for groups of users*. In *ECSCW 2001*, pp. 199–218. Springer. 23

Olay (2020). *Skin Advisor*. Online: https://www.olay.com/en-us/skinadvisor; accessed 12-05-2022. 27

O'Reilly III, C. A. (1982). *Variations in decision makers' use of information sources: The impact of quality and accessibility of information*. Academy of Management journal, 25(4):756–771. 1

Orlikowski, W. J. (2002). *Knowing in practice: Enacting a collective capability in distributed organizing*. Organization science, 13(3):249–273. 12

Ortega, F., Bobadilla, J., Hernando, A. & Gutiérrez, A. (2013). *Incorporating group recommendations to recommender systems: Alternatives and performance*. Information Processing & Management, 49(4):895–901. 23

Pantano, E. (2014). *Innovation drivers in retail industry*. International Journal of Information Management, 34(3):344–350. 18

Papadimitriou, A., Symeonidis, P. & Manolopoulos, Y. (2012). *A generalized taxonomy of explanations styles for traditional and social recommender systems*. Data Mining and Knowledge Discovery, 24(3):555–583. 22

Park, D. H., Kim, H. K., Choi, I. Y. & Kim, J. K. (2012). *A literature review and classification of recommender systems research*. Expert systems with applications, 39(11):10059–10072. 22

Park, M.-H., Park, H.-S. & Cho, S.-B. (2008). *Restaurant recommendation for group of people in mobile environments using probabilistic multi-criteria decision making*. In *Asia-pacific conference on computer human interaction*, pp. 114–122. Springer. 22

Park, Y. & Gretzel, U. (2010). *Influence of consumers' online decision-making style on comparison shopping proneness and perceived usefulness of comparison shopping tools*. 25, 33

Payne, J. W., Bettman, J. R. & Johnson, E. J. (1992). *Behavioral Decision Research: A Constructive Processing Perspective*. Annual Review of Psychology, 43(1):87–131. 26

Payne, J. W., Samper, A., Bettman, J. R. & Luce, M. F. (2008). *Boundary conditions on unconscious thought in complex decision making*. Psychological Science, 19(11):1118–1123. 7

Poushneh, A. & Vasquez-Parraga, A. Z. (2017). *Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy*. Journal of Retailing and Consumer Services, 34:229–234. 18, 54

Punj, G. (2012). *Consumer decision making on the web: A theoretical analysis and research guidelines*. Psychology & Marketing, 29(10):791–803. 15

Quijano-Sanchez, L., Recio-Garcia, J. A. & Diaz-Agudo, B. (2011). *Happymovie: A facebook application for recommending movies to groups*. In *2011 IEEE 23rd international conference on tools with artificial intelligence*, pp. 239–244. IEEE. 22

von Reischach, F., Guinard, D., Michahelles, F. & Fleisch, E. (2009). *A mobile product recommendation system interacting with tagged products*. In *2009 IEEE International Conference on Pervasive Computing and Communications*. IEEE. 26

Ricci, F., Rokach, L., Shapira, B. & Kantor, P. B. (eds.) (2011). *Recommender Systems Handbook*. Springer US. 1, 19, 25, 26

Rippé, C. B., Weisfeld-Spolter, S., Yurova, Y. & Sussan, F. (2015). *Is there a global multichannel consumer?*. International Marketing Review. 18, 33

Romiszowski, A. & Mason, R. (2013). *Computer-mediated communication*. In *Handbook of research on educational communications and technology*, pp. 402–436. Routledge. 1, 13, 31

Sae-Ueng, S., Pinyapong, S., Ogino, A. & Kato, T. (2008). *Personalized shopping assistance service at ubiquitous shop space*. In *22nd International Conference on Advanced Information Networking and Applications-Workshops (aina workshops 2008)*, pp. 838–843. IEEE. 26

Sager, K. L. & Gastil, J. (1999). *Reaching consensus on consensus: A study of the relationships between individual decision-making styles and use of the consensus decision rule*. Communication Quarterly, 47(1):67–79. 10

Sager, K. L. & Gastil, J. (2006). *The origins and consequences of consensus decision making: A test of the social consensus model*. Southern Communication Journal, 71(1):1–24. 11, 30

Santos, S. & Gonçalves, H. M. (2021). *Touchpoints and Channels: Classifications, Characteristics, and Issues for Future Research*. In *Marketing and Smart Technologies*, pp. 311–323. Springer. 13

Schafer, J. B., Konstan, J. A. & Riedl, J. (2001). *E-commerce recommendation applications*. Data mining and knowledge discovery, 5(1):115–153. 25, 26, 33

Schmidt, J. B. & Spreng, R. A. (1996). *A proposed model of external consumer information search*. Journal of the academy of Marketing Science, 24(3):246–256. 17

Schoenfeld, A. H. (2010). *How we think: A theory of goal-oriented decision making and its educational applications*. Routledge. 5

Shufford, K. N., Hall, D. L., Randall, A. K., Braunstein, B. M., O'Brien, M. M. & Mickelson, K. D. (2021). *Connected while apart: Associations between social distancing, computer-mediated communication frequency, and positive affect during the early phases of COVID-19*. Journal of Social and Personal Relationships, 38(10):2906–2920. 1, 12

Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason*, vol. 3. MIT press. 1, 6

Smink, A. R., Frowijn, S., van Reijmersdal, E. A., van Noort, G. & Neijens, P. C. (2019). *Try online before you buy: How does shopping with augmented reality affect brand responses and personal data disclosure*. Electronic Commerce Research and Applications, 35:100854. 27

Solomon, M. R., Dahl, D. W., White, K., Zaichkowsky, J. L. & Polegato, R. (2014). *Consumer behavior: Buying, having, and being*, vol. 10. Pearson Toronto, Canada. 15

Spreer, P. & Kallweit, K. (2014). *Augmented reality in retail: assessing the acceptance and potential for multimedia product presentation at the PoS*. Transactions on Marketing Research, 1(1):20–35. 18

Sproull, L., Kiesler, S. & Kiesler, S. B. (1991). *Connections: New ways of working in the networked organization*. MIT press. 13

Streufert, S. C. (1973). *Effects of information relevance on decision making in complex environments*. Memory & Cognition, 1(3):224–228. 2, 7, 8

Taylor, E. (2016). *Mobile payment technologies in retail: a review of potential benefits and risks*. International Journal of Retail & Distribution Management. 17, 34

Torres-Ruiz, M., Mata, F., Zagal, R., Guzmán, G., Quintero, R. & Moreno-Ibarra, M. (2020). *A recommender system to generate museum itineraries applying augmented reality and social-sensor mining techniques*. Virtual Reality, 24(1):175–189. 27

Välkkynen, P., Boyer, A., Urhemaa, T. & Nieminen, R. (2011). *Mobile augmented reality for retail environments*. In *Proceedings of Workshop on Mobile Interaction in Retail Environments in Conjunction with MobileHCI*. 18

Vroman, K. & Kovacich, J. (2002). *Computer-mediated interdisciplinary teams: Theory and reality*. Journal of Interprofessional Care, 16(2):159–170. 13

Wald, A. (1950). *Statistical decision functions*. 5

Wang, Y., Wang, Y., Patel, S. & Patel, D. (2006). *A layered reference model of the brain (LRMB)*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 36(2):124–133. 5

Wolny, J. & Charoensuksai, N. (2014). *Mapping customer journeys in multichannel decision-making*. Journal of Direct, Data and Digital Marketing Practice, 15(4):317–326. 2, 8, 15, 16, 17, 18, 33

Wood, R., Bandura, A. & Bailey, T. (1990). *Mechanisms governing organizational performance in complex decision-making environments*. Organizational behavior and human decision processes, 46(2):181–201. 7

Yu, Z., Zhou, X., Hao, Y. & Gu, J. (2006). *TV program recommendation for multiple viewers based on user profile merging*. User modeling and user-adapted interaction, 16(1):63–82. 22

Zaichkowsky, J. L. (1985). *Measuring the involvement construct*. Journal of consumer research, 12(3):341–352. 15, 33

Zakay, D. (1993). *The impact of time perception processes on decision making under time stress*. In *Time pressure and stress in human judgment and decision making*, pp. 59–72. Springer. 11

Zimmermann, R., Mora, D., Cirqueira, D., Helfert, M., Bezbradica, M., Werth, D., Weitzl, W. J., Riedl, R. & Auinger, A. (2022). *Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence*. Journal of Research in Interactive Marketing, ahead-of-print. 27

Zopounidis, C. & Pardalos, P. M. (2010). *Handbook of multicriteria analysis*, vol. 103. Springer Science & Business Media. 5

Appendix
Papers included in this Dissertation

Paper I. Preference Elicitation and Negotiation in a Group Recommender System

Reproduced with permission from Springer Nature: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Preference Elicitation and Negotiation in a Group Recommender System*. In *Human-Computer Interaction – INTERACT 2015*, Lecture Notes in Computer Science, pp. 20–37.

https://doi.org/10.1007/978-3-319-22668-2_2

Preference Elicitation and Negotiation in a Group Recommender System

Jesús Omar Álvarez Márquez(✉) and Jürgen Ziegler

University of Duisburg-Essen, Duisburg, Germany
{jesus.alvarez-marquez, juergen.ziegler}@uni-due.de

Abstract. We present a novel approach to group recommender systems that better takes into account the social interaction in a group when formulating, discussing and negotiating the features of the item to be jointly selected. Our approach provides discussion support in a collaborative preference elicitation and negotiation process. Individual preferences are continuously aggregated and immediate feedback of the resulting recommendations is provided. We also support the last stage in the decision process when users collectively select the final item from the recommendation set. The prototype hotel recommender Hootle is developed following these concepts and tested in a user study. The results indicate a higher overall satisfaction with the system as well as a higher perceived recommendation quality when compared against a system version where no negotiation was possible. However, they also indicate that the negotiation-based approach may be more suitable for smaller groups, an aspect that will require further research.

Keywords: Group recommender system · Group preference elicitation · Negotiation · Decision making

1 Introduction

Over the recent years, recommender systems have proven beneficial in supporting users when selecting or buying items from large sets of alternatives [30]. Buying something in a virtual shop, deciding which film to watch or planning where to go on holidays can easily become a tedious task when solely relying on manual search and filtering techniques, which may lead to information overload and choice difficulties. Therefore, the importance of recommender systems has increased fast in the last years, being now used widely throughout the internet. While the field of recommendations for single users has already been deeply explored, the same cannot be said about group recommender systems. Even though a significant number of group recommenders have been developed in the past years [5, 18], there is still a range of issues which have not been sufficiently investigated so far.

Most group recommending approaches rely on existing user profiles which are either aggregated into a single group profile (model aggregation) before generating group recommendations, or which are used for calculating individual recommendations that are subsequently aggregated, using a variety of different strategies (recommendation aggregation). However, while sufficient profile information is often not available

J. Abascal et al. (Eds.): INTERACT 2015, Part II, LNCS 9297, pp. 20–37, 2015.
DOI: 10.1007/978-3-319-22668-2_2

in the case of single users – either due to a cold start condition, or because users do not want their profile to be stored – this problem is even more pertinent for groups where the likelihood of each user having a stored profile that can be exploited by the recommender is relatively low. This is especially the case for ad hoc groups who gather spontaneously or who come from different organizational contexts. A further issue is the situational variability of the group members' preferences. This is also a problem in single-user recommending, but is aggravated by the fact that the inherent heterogeneity of preferences in a group may be amplified due to different responses to the situational context. These issues ask for methods that can elicit group preferences on the fly and that can aggregate individual preferences in a manner that best suits the individual users as well as the group as a whole.

Solving the complex trade-off between the degree of satisfaction of individual users and the group as a whole is typically attempted by applying one out of set of fixed strategies, such as averaging the satisfaction of all group members or minimizing discomfort for the least satisfied user. However, fixed strategies do not take the dynamics of group settings and situational needs into account. In particular, the aspect of social interaction when moving towards a joint decision is typically not sufficiently supported in existing group recommenders.

In this paper, we propose a novel method that tries to approach group recommendations from the point of intersection of traditional group recommenders and group decision making theory, allowing users to collaboratively create a preference model (thus addressing collaborative preference elicitation [28]), from which recommendations are generated. In this process, group interaction can happen at two (tightly intertwined) stages: (1) users can online discuss and negotiate preferences stated by others, and (2) they can discuss and rate items taken from the recommendation set to arrive at a final consensus decision.

Following the idea that computer-mediated discussion groups have more equal member participation [32], the goal is to avoid unfair situations in which some users might not be satisfied with the items proposed by the system. Our system supports remote online negotiation, although the approach can also be adapted to co-located settings. Each user can specify an individual preference model by freely adding desired features, using an explicit preference elicitation approach [27]. The individual preferences are then aggregated to form the group preference model and to determine an initial set of recommendations. All members' preferences, as well as the group aggregation, are visible to the participants. Most importantly, individual preferences can then be negotiated in a system-supported manner: by group discussion, members may thus be able to convince other users to modify their preferences, so the group model changes to better match all members' desires. Recommendations are continuously calculated and updated when the group preferences change, thus allowing users to immediately see the effect of their actions. Different mechanisms are provided for discussing and reaching an agreement, both for the creation of a group preference model and for the final item selection.

In the following, we first survey related research before presenting the conceptual aspects of our approach. We then describe the prototype implementation *Hootle* and its user interface design. We report on a user study we performed with groups of different sizes and conclude by summarizing our work and outlining future work.

2 Related Work

While the field of recommending items for single users has already received a great deal of attention in recent research, leading to quite effective recommendation methods, recommender systems for groups are, in comparison, a still less deeply investigated area. Various group recommender systems have been developed over the recent years, starting from early systems such as *MusicFX* [19], a group music recommender, that use different approaches for generating recommendations [5, 12]. However, there are still many open research questions concerning, for example, the best approach to aggregating individual preferences, techniques for responding to the situational needs of the group, or supporting the social interaction processes in the group for converging on a joint decision.

To structure the wide range of different aspects involved in group recommending, [14] suggest a design space comprising the dimensions preference input (including dynamic aspects), process characteristics, group characteristics, and (presentation of) output. In the process dimension, an important aspect is how individual, possibly conflicting preferences can be merged to obtain recommendations that best fit the group as a whole. Although different approaches in group recommenders gather and represent users' preferences in different ways, they commonly use one of two schemas [12]:

Aggregation of Item Predictions for Individual Users (Prediction Aggregation). This approach assumes that for each item, it is possible to calculate a user's satisfaction, given the user's profile. Then, using the calculated predictions and making use of some specific aggregation strategy, items are sorted by the group's overall satisfaction. In [9] a video recommender that uses this strategy is described; also *Polylens* [26], a system that suggests movies to small groups of people with similar interests, based on the personal five-star scale ratings from *Movielens* [8] uses this method.

Construction of Group Preference Models (Model Aggregation). Instead of predicting matching items for each user, the system uses information about individual members to create a preference model for the group as a whole. Recommendations are generated by determining those items that best match the group model. The number of possible methods for creating the group's model is even bigger than it is for prediction aggregation strategies. For example, in *Let's Browse* [15] the group preference model can be seen as an aggregation of individual preference models. In *Intrigue* [1, 2] (which recommends sightseeing destinations for heterogeneous groups of tourists) the group preference model is constructed by aggregating preference models of homogeneous subgroups within the main group. *MusicFX* [19] chooses background music in a fitness center to accommodate members' preferences, also by merging their individual models. AGReMo [4] recommends movies to watch in cinemas close to a location for ad hoc groups of users, creating the group's preference model not only by individual model aggregation but also taking into account some specific group variables (e.g. time, weight of each member's vote). Furthermore, the *Travel Decision Forum* [10, 11] creates a group preference model that can be discussed and modified by the members themselves, aiming to non-collocated groups who are not able to meet face to face, allowing asynchronous communication.

Regardless of whether the aggregation is made before or after generating recommendations, an aggregation method that is appropriate for the specific group characteristics needs to be chosen. There are a number of voting strategies, empirically evaluated in [18], that have been used in actual group recommender systems. Some typical strategies (and systems using it) are:

- **Average strategy**, where the group score for an item is the average rating over all individuals (*Intrigue, Travel Decision Forum*).
- **Least misery strategy**, which scores items depending on the minimal rating it has among group members (*Polylens, AGReMo*).
- **Average without misery strategy**, consisting in rating items using an average function, but discarding those where the user score is under a threshold (*MusicFX, CATS* [20–23]).
- **Median strategy**, which uses the middle value of the group members' ratings (*Travel Decision Forum*).

On another dimension, the question of preference elicitation has to be solved, which is concerned with how the user-specific preference information needed to generate recommendations is obtained. One approach is to let users rate a number of items in advance and to derive preferences from this set of ratings. *AGReMo*, for instance, requires group members to create their own model of individual preferences before the group meeting takes place by rating movies that they already saw. In *Travel Decision Forum* each participant starts with an empty preference form that has to be filled with the desired options, so group members define new preferences for each session. A more interactive approach, although for single user systems, is described in [17] which requires users to repeatedly choose between sets of sample items that are selected based on latent factors of a rating matrix. The techniques mentioned also address the cold-start problem when no user profile is available up-front but initially require some effort on the part of the user to develop a sufficiently detailed profile.

However, most preference elicitation techniques do not take group interaction into account. As pointed out in [16], to obtain adequate group recommendations it is not only necessary to model users' individual preferences, but also to understand how a decision among group members is reached. While research on group decision making [31] is concerned with collaboratively making choices, focusing on the social process and the outcome, these aspects have mostly not been addressed in the development of group recommender systems. The process of group decision making involves a variety of aspects, such as the discussion and evaluation of others' ideas, conflict resolution, and evaluating the different options that have been elaborated. Also interesting for our research is the concept of consensus decision-making [7], which seeks for an acceptable resolution for the whole group. Within this context, Group Decision Support Systems (GDSS) have emerged, that aim at supporting the various aspects of decision making [24, 25]. Only few recommender systems attempt to include aspects of group decision theory, for instance, by introducing automated negotiation agents that simulate discussions between members to generate group recommendations [3]. However, supporting the entire preference elicitation and negotiation process that may occur when users take recommender-supported decisions is, to our knowledge, not realized by current group recommenders.

Lastly, taking into account the social factor that is involved in group recommendation, one needs to contemplate the question whether a user would be willing to change personal preferences in favor of the group's desires, bringing up the importance of group negotiation. Again in the *Travel Decision Forum*, users are able to explore other members' preferences, with the possibility to copy them or propose modifications. The *Collaborative Advisory Travel System* (*CATS)* focuses on collocated groups of persons gathered around a multi-touch table. Recommendations are made by collecting critiques (users' feedbacks respecting recommended destinations) that can be discussed face to face, since the system gives visual support to enhance awareness of each other's preferences. The main difference between CATS and the system proposed here is that the former is focused in critiquing items once they have been recommended, while the latter allows negotiation already in the preference elicitation stage.

3 Preference Elicitation and Negotiation Method

The method developed involves an iterative process of specifying, discussing and negotiating preferences in a remote collaboration setting. Instead of only discussing recommendations produced based on user profiles, interaction among group members is supported right from the beginning of the preference elicitation process. The overall process comprises the following stages which are not meant as sequential steps but which can basically be performed in any order (algorithmic and interface details are described in the next chapter):

1. Users begin by selecting desired features from a set of attributes describing the items available. Since the feature sets may be very large (e.g. cities in our example hotel recommender, users can first search for the features they want and place them in a private area).
2. By moving a feature to the user's individual preference list, the feature becomes active and is visible to other group members. Several features can be placed and rank ordered according to the relevance they have for the user.
3. The individual feature lists are constantly aggregated in a common, ranked group preference list and the recommendations that best match the current group model are immediately generated and shown to the group.
4. Users can discuss preferences stated by others and negotiate them by using a 'petition' function, potentially trading in own preferences for features other users want. Based on the discussions and negotiations, users may change their preferences which is again immediately reflected in the group model and the resulting recommendations.
5. From the recommendations users can at any time select the item(s) they really like and propose them to the other participants who can accept them or propose alternatives. Also in this stage of the process, discussions are supported by the system.

The closed loop interaction with immediate feedback in the group model and the recommendations increases participants' awareness of others' preferences and the effects their own preference changes have on the group results. The approach also entails aspects of critique-based recommenders since users can criticize or accept proposed features or recommended items. In contrast to fully automated recommender system, users have a higher level of control over the process and can easily adapt it to their current situational needs and context.

4 Description of the System

To demonstrate our approach we designed and implemented a prototype group recommender system that employs content-based techniques. The system is in principle applicable in a wide range of application areas, such as candidate selection, requirements specification, or leisure activities, as long as it is possible to obtain the properties of the items to be recommended. For demonstration purposes, we chose hotel selection for group travel as application area and use an Expedia dataset consisting of 151.000 hotel entries with descriptive information.

Figure 1 shows a screenshot of the user interface, described as following:

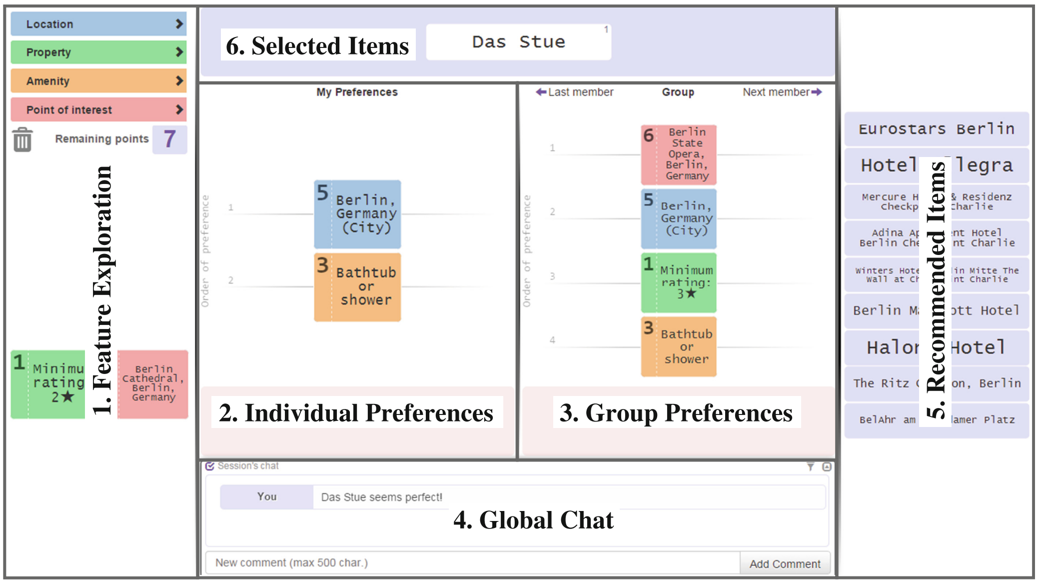


Fig. 1. Areas of the interface.

1. **Feature exploration.** This area consists of a set of defined filters that let users search for specific attributes and a space to store the selected ones. For example, filters could be location, facilities or nearby points of interest.
2. **Individual preferences.** Features selected in area 1 can be added here by drag-and-drop, meaning that the user wants these features to be present (or excluded) in the recommended items (more details in the section about Individual Preferences). Users can also rank their preferences to express different levels of importance.

3. **Group preferences**. A ranked aggregation of all individual preferences is displayed in this area. It is also possible for users to navigate through the preferences of other participants here.
4. **Global chat.** In this section, the group can discuss arbitrary questions that come up in the decision process. Requests for preference changes ("petitions") and comments about specific features can also be displayed here.
5. **Recommended items.** Here, the items that best match the current group preferences and their relative weight are shown. The list is constantly updated in real-time when users add or change features.
6. **Recommendations selected by users.** From the recommendations area, users can pick the items they like most, and place them here. This space works as a shared area, so each item added here is visible to all participants.

4.1 Feature-Based Preference Elicitation

Individual preferences are defined by each group member by selecting features from the exploration area, where they can use different filters to locate them. Later, features can be placed into the user's individual preference space. The system allows to specify both positive and negative features.

Positive features. Apositive property means that a user wants it to be found in the recommendations. Users can specify an order of preference among positive attributes by dragging them to a higher or lower position in the list, which denotes the degree of importance that the user gives to each feature. Multiple features may have the same preference level.

Negative features. Negative properties are those that the user does not want to get as feature of the recommended items. They are placed inside a subspace within the

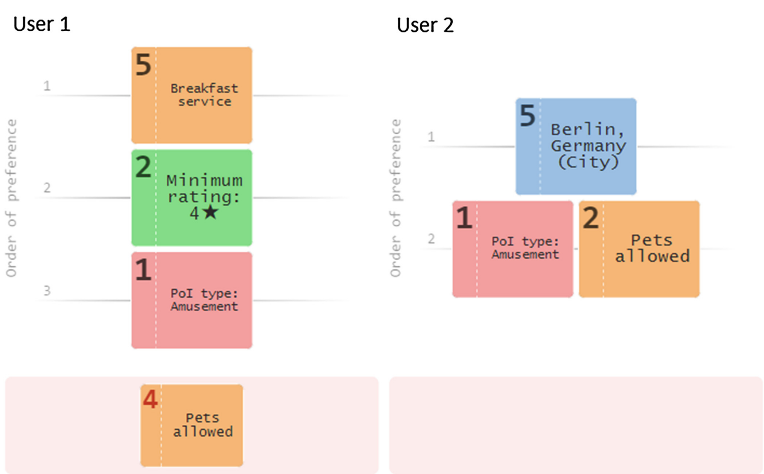


Fig. 2. Example of preference areas belonging to two different users. The ordered list represents the positive (desired) attributes, while the area at the bottom contains the negative (vetoed) ones. The cost of each attribute can be found at the top-left corner.

individual area (Fig. 2), called the veto area. Vetoed attributes have no preference order.

Cost of features. When users specify a large number of features as preferences, several problems may arise: first, it may be difficult to create meaningful integrated group preferences because the probability that features contradict each other increases, requiring more complex and longer negotiation processes. Second, users may over-specify their preferences making it difficult or impossible to calculate well-matching recommendations. We therefore decided to devise a mechanism that gently pushes users towards only specifying the features they really want.

For this purpose, a method for measuring the cost of each feature has been implemented. Each attribute has a related cost depending on how restrictive it is (i.e. how many items are left after using it as filter over the database). When a user selects a feature he or she pays for it from a limited budget. Users only have a number of tokens to exchange for attributes so they have to choose which ones are most important. This way, users selecting very restrictive features will only be able to create a small list of preferences as they will cost more tokens. It is also necessary to remark that the cost for positive attributes differs from the one for negatives. Positive attributes are more expensive the more restrictive they are; for negative features, more restrictiveness means less cost.

Group Preferences. While creating their individual preference lists, users can immediately see the overall results for the group. Inside the group preference area, an aggregation of all individual user preferences is displayed. This list is called the group preference list. The aggregation of individual preferences is performed using a variant of the Borda Count method, combined with rules regarding the vetoed attributes.

Borda Count is a voting method in which voters rank options or candidates in order of preference. In standard Borda Count, each option receives a score depending on its rank, and to obtain the aggregated score the points that each voter has given to it are summed up. In the case at hand, not only the rank of each option has been taken into account, but also its cost. When a user chooses to place a relatively expensive (restrictive) feature in the individual preference list, it is fair to think that the user cares more about this specific attribute. The equation used to calculate the aggregated score of an attribute $\boldsymbol{i}$ is presented in (1), where $\boldsymbol{u}$ is the number of group's members, $\boldsymbol{n}$ is the total number of different attributes used, $\boldsymbol{p_{ij}}$ is the preference value given to the attribute i by the user j, $\boldsymbol{c_i}$ is the cost of the attribute i and $\boldsymbol{\lambda}$ is used to correct the importance of the cost (with $\lambda = 0$ the result would be a standard Borda Count voting aggregation).

$$PAtt_i = \sum\nolimits_{j=0..u} \left(\frac{1}{n} \left(n - p_{ij} \right) \right) + \frac{\lambda c_i}{n} \tag{1}$$

Attributes only receive points if users include them in their preferences. Finally, the group preference list is created by calculating the total score for each item and sorting them as usual (Fig. 3).

Vetoing a feature is a strong statement, it means that the person who stated it really does not want items with this feature. It would be desirable to avoid this feature, even if

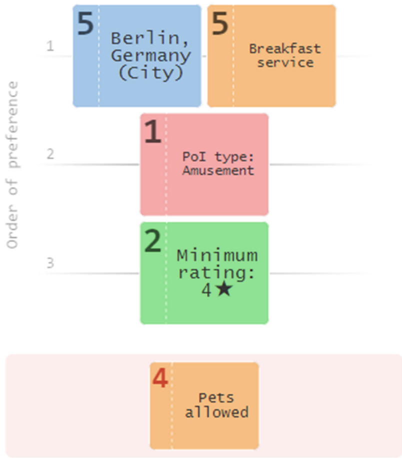


Fig. 3. Resulting preference setting for the group, using the individual lists shown at Fig. 2.

someone else in the group still wants it. Thus, vetoed attributes are removed from the group preference list and will not appear in any of the recommendations.

4.2 Generating Recommendations

Based on the aggregated user preferences the system applies a content-based filtering method to generate recommendations (Fig. 4). In content-based filtering, items are described by a set of attributes, and each user has a profile of preferences indicating the item properties the user likes. In our case, the individual preference set in a session represents the full user profile, thus, the system is applicable in cold-start situations where no user profile exists yet.

To generate recommendations, group preferences are compared to the items' properties in order to find the best matching ones. First, the system removes all the items

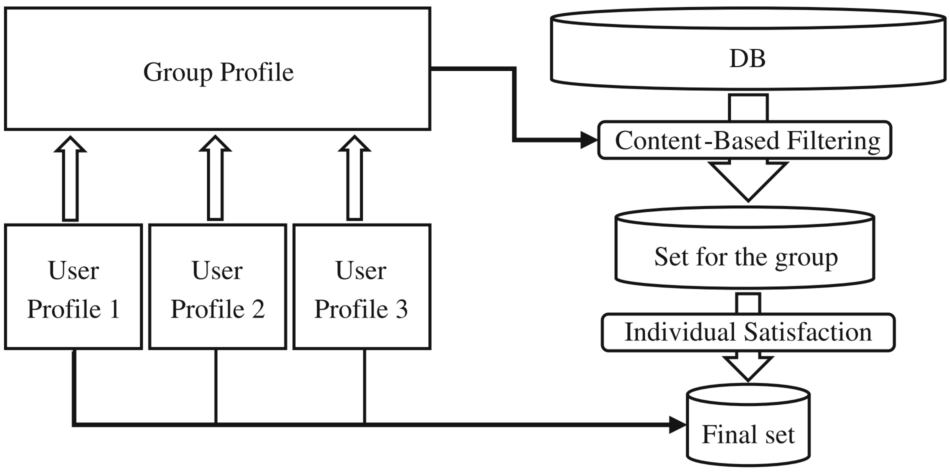


Fig. 4. Scheme of the filtering process.

that contain a vetoed attribute. The remaining items receive a score based on how many positive features they match, their total score being the sum of their attributes' values. The value of each attribute comes defined by the Borda Count method previously described, so attributes with higher preference levels will give higher score values to the items containing them. For distance attributes (coordinates, regions or points of interest), the value they were assigned by the Borda Count is modified depending on how far an item is from the given feature (closer items obtain higher scores).

If the system would simply present the ten top scored items, it could happen that for some users whose attributes are lower in the group preference list, no good options are returned. Since the main purpose of the system is to provide a negotiation environment, it seems necessary to return a well-balanced set of items, in terms of member satisfaction. For this reason, a subset of items is extracted, within the already found, in a way that for each user there is at least one acceptable option, but giving at the same time importance to the items that satisfy the group as a whole. An item is considered acceptable for a participant when his/her satisfaction level concerning this option is higher than a given threshold. Satisfaction is calculated taking into account the individual preference model defined by a user, in a similar way an item's group score is calculated, but divided by the maximum points an item could receive (that is, when an item contains all the features a user wants). Finally, the selected items are presented to participants in the recommendation area of the screen (5 - Recommended Items in Fig. 1).

As said before, the system is applicable without requiring the prior availability of stored user profiles which is particularly beneficial in group contexts for the reasons mentioned earlier. However, in principle more complex and longer-term user profiles could be built if past choices were saved for future sessions. If this option was used and is acceptable for users, the interaction effort needed for specifying the desired features could be reduced, just specifying changes in the existing profile, and possibly increasing the precision of the recommendations.

4.3 Negotiation

User preferences are typically not a static phenomenon but are influenced by the situational context of the group and the social interaction that takes place within it. Users may also differ in the extent to which they have already formed their objectives at the beginning of the group process. They may react to preferences expressed by others, either accepting or rejecting them. They may also be willing to dispense with a desired feature if someone else in the group accepts one of their other preferences, thus embarking on a negotiation process with other group members. For these reasons, our system provides several functions that specifically support discussion, negotiation and consensus finding among group members.

Communication. Users need the possibility to express their opinions about the decision process as a whole as well as about specific preferences stated by others. To support these types of communication, two methods are implemented in the system.

Discussion threads and global chat. Each feature has its own discussion thread, which means that users can access it and say what they think about a specific property, keeping the comments organized by attribute. A global chat is also available, placed in area 4 displayed in Fig. 1. The global chat lets participants talk about arbitrary aspects of the current session, and also informs group members about recent updates in specific comment threads.

Petitions. Petitions are requests such as removing a feature or changing its rank. It is not possible to request the addition of an attribute, as adding a feature to one's individual list is already an implicit petition to the rest of the group: every user wants the others to adopt the same preferences as he/she has, since this would increase the fit of the recommendations with this user's wishes.

Finding and Resolving Conflicts. Conflicts appear when two or more participants want features that contradict each other. Several mechanism help to resolve such situations. First, users can explore the individual preferences of other participants and discuss them if a conflict occurs.

Second, once a set of recommendations is presented, users can access information about each item recommended. Also, those entries in the group preference list that are not fulfilled by an item are highlighted in that list. Thus, when a user likes a recommendation, he/she can see the preferences that are in conflict with it and try to change the opinion of the members who added them.

Finally, for each recommendation, the calculated grade of satisfaction of each user can be displayed in a spider diagram, so the group may choose items that are more balanced with respect to the members' individual desires (i.e. are less conflictive).

Proposing Items. From the recommendation area, users are able to express their approval for a specific recommended item by placing it into the "recommendations selected by users" space (area 6 in Fig. 1). This step shows the group that one user likes a recommendation and proposes it as option. The other participants now can accept it as a good option, reject it or just ignore it, waiting for more proposals to show up.

4.4 Repeat and Decide

The *"adding features-get recommendations-negotiate"* cycle can be repeated several times, narrowing down the recommendations given with each new iteration, until the group reaches agreement. If and when consensus is reached, however, is something that only the group itself is able to decide. As has been said in the previous section, users can add items that they like into a shared area, so the others can express their acceptance about it. For some groups, the item to be finally selected may be the one that is accepted by more than fifty percent of the members; in other cases, there may be situations where all users have accepted an item except one who finds it unsatisfactory. While a fixed group recommendation strategy, for example, a 'least misery' approach that might seem applicable in the latter case, would always try to satisfy user needs in one prescribed manner, we believe that the system cannot generally resolve such decision problems. Although the system provides tools for preference specification, discussion and acceptance measuring, it is up to the users to decide whether a recommendation fits their needs or not and to make the final choice.

5 Evaluation

To evaluate our approach, we performed a user study with several groups comprising between three and five users. We did not consider larger groups at this point because we believe this group size to be typical for the application domain chosen which is selecting a hotel for a joint leisure or business trip. Also, Hootle, our Web-based prototype implementation of the approach, while still work in progress, is stable enough to support this group size but still has to be tested for larger-scale trials. The main objectives of this study were to determine the usability of the approach and the quality of the resulting recommendations, as well as, more specifically, to analyze the impact of the cooperative preference elicitation and negotiation tools developed.

5.1 Setting and Experimental Tasks

To assess whether the preference elicitation, negotiation and recommendation methods developed benefit group decision processes, we tested two different versions of the system where one served as baseline for comparison. While one system version provided the full set of functions described including group discussion support (hereafter version D – Discussion), we restricted the second version to specifying preferences and calculating recommendations (version ND – No Discussion), similar to a conventional group recommender system, but still offering the possibility to specify preferences in an ad hoc manner without using existing user profiles. We decided against using an existing alternative group recommender for comparison because the systems would have differed in too many aspects, making it difficult to pinpoint the specific benefits of the proposed innovations. In both cases, we make use of a hotel database provided by Expedia with 151,000 entries. For each hotel, a full description and a set of attributes, including property and room amenities (within a total of 360 possibilities), locations (258,426) and points of interest nearby (94,512) was available. We deliberately decided to focus the negotiation and decision process on the objective properties of the items, excluding price information which would have opened up additional questions concerning economic concerns and behavior in the test groups. This aspect, however, will be subject of future research.

We prepared two types of task scenarios with different levels of complexity:

- In an 'introductory' task, the group was instructed to select a hotel knowing beforehand some common, desired attributes, as well as the location of the hotel. This task also served as a training session for the application, to allow participants to explore the functions and possibilities the system supplies. Two scenarios for this task were presented:
  - Your group will be participating at a conference in Berlin. As the conference always provides lunch and dinner, you just need to find a hotel including breakfast. Your conference takes place near the Brandenburg Gate.

  - Your group wants to enjoy some days on the beach. You already decided to go to an apartment, as you want to prepare meals on your own. Everyone loves Spain so you also decided to go to Marbella.
- In the 'open' task which was always performed after the introductory task, only unspecific instructions were given to the group such as "Find a place to stay during summer vacation". The possible scenarios were:
  - It is summertime. You and your friends really need to get out of the daily routine. Discuss where to stay.
  - Your group wants to do some kind of city trip. Where are you going to?

To avoid the problem that in a test situation, participants do not bring with them the objectives and preferences they would have in a real-life decision situation, or might comply too quickly with the wishes of other participants, we tried to artificially induce different backgrounds and objectives for each group member. For this purpose, we created a set of role cards for the second task, depending on the scenario used. With this method, we expected to generate conflicts and discussion when randomly distributing the role cards among group's members. As an example, the role cards for the first scenario in task 2 were (abbreviated here):

1. You're a sport addict. You like to eat healthy and don't trust in hotel food. You hate giant hotels and prefer small pensions or camping sites.
2. You're allergic to nearly everything. Vacation at a camping site would be like a death sentence to you. You prefer the pool over the sea. You don't want to do anything so you prefer all inclusive.
3. You like to go for long hikes. You're fascinated by mountains. You don't want to cook but you won't be there during the day so you just need breakfast and dinner.
4. You're into cultural things. If you go on vacation, you want to see things. You also like to go out for dinner so breakfast only would totally fit your needs.
5. You like to party. As you won't be able to prepare your own food, there should be someone who helps you with this. More important is the location of your hotel. Nobody wants to walk for an eternity to go clubbing.

5.2 Method

A total of 48 students were recruited as participants (5 male, 43 female, average age of 20.94, σ 5.018), distributed in groups of different sizes: 4 groups of 3 persons (12), 4 groups of 4 persons (16) and 4 groups of 5 persons (20). Two groups of each size ran a full version of the system (D), while the other two groups tested the version without negotiation support (ND). Since the system is Web-based, all users were provided with a normal desktop computer with a display screen of 21 in and running the same browser. They sat in a large lab room but were separated from each other and instructed to only communicate via the means provided by the system.

Each group first received a brief introduction to the system which was dependent on whether the negotiation support was turned on or off for the group. After a brief trial, they were asked to work on the two decision tasks, always in the order introductory task – open task. Before beginning the second task, they all received randomly one of the role cards.

For the groups using version D, a task was considered complete when they reached consensus about their preferred hotel or when they decided that it was not possible to find agreement. Since the groups with version ND were not able to communicate at all, their job consisted in defining their own preference model and, when the whole group had done this, each user separately selected a hotel from the resulting set of recommendations.

The first task including the explanation of the system was limited to a maximum of 40 min. As the explanation was no longer necessary, the second task, although more complex, should also be completed during this time.

After completing both tasks, participants were asked to fill in a questionnaire regarding aspects such as the quality of the recommendations or the ease-of-use of the system, using a 1-5 scale. The questionnaire comprised the SUS items [6] to compare the system against a well-established baseline as well as items from two recommender-specific assessment instruments (User experience of recommender systems [13] and *ResQue* [29]). The recommender-specific items were measuring mainly the constructs *user-perceived recommendation quality*, *perceived system effectiveness, interface adequacy,* and *ease of use*.

5.3 Results and Discussion

All tasks were finished within the allotted time. The D an ND groups differ on a considerable number of criteria. The members in ND groups were not able to choose the same hotel in a single instance. In two of these cases, some users couldn't even find a hotel that they liked when realizing the open task. On the other hand, all groups with version D were able to choose one unique hotel in both tasks, despite starting the

Table 1. Results of the questionnaire (all the D/ND differences $p > 0.05$, effects of group size were significant).

System version		No discussion				Discussion			
Group Size		3	4	5	Avg.	3	4	5	Avg.
Overall satisfaction	m	3.40	3.00	3.70	3.39	4.33	4.00	3.60	3.92
	σ	0.54	1.20	0.48	0.83	0.51	0.53	0.96	0.77
Would recommend it	m	3.20	2.38	3.30	2.96	3.50	3.25	3.30	3.33
	σ	1.30	1.06	0.67	1.02	0.83	0.70	1.06	0.86
Would use it again	m	2.40	2.50	3.10	2.74	3.17	3.13	3.00	3.08
	σ	0.89	0.92	1.10	1.01	0.75	0.99	0.66	0.77
Would use it frequently	m	1.60	1.88	2.30	2.00	2.67	2.75	2.70	2.71
	σ	0.54	0.64	0.67	0.67	0.81	1.04	0.94	0.90
Recommendations were well chosen	m	3.20	3.38	3.80	3.52	4.33	3.38	4.00	3.88
	σ	0.83	0.74	0.78	0.79	0.51	0.74	0.47	0.68

process with strongly different individual preferences. To achieve this joint decision, users had to iterate several times through the *"adding features-get recommendations-negotiate"* cycle, as well as to renounce some desired features due to the influence exerted by other members through discussions and petitions.

In terms of overall usability, both system versions received a SUS score which can be considered as borderline good with no differences between the two systems (ND = 68, D = 69). We performed a 2 × 3 ANOVA with system version and group size as independent variables and questionnaire item scores as dependent variables. Most item responses did not show significant differences between the two system versions which may be due to the limited number of groups tested. In Table 1, we list some of the results that were significant at a .05 level. Users in the discussion condition were overall more satisfied with the system, are more likely to recommend it to others and would be willing to use the system again and also more frequently. Also, the accuracy of the recommendations was rated higher in the discussion groups. While these results speak in favour of the discussion version, there appears to be an interesting interaction effect between system versions and group size. Generally, satisfaction and willingness to use and recommend the system tend to be higher for the small groups than the large groups when discussion is available. Concerning recommendation quality, the largest group had the highest ratings in the no-discussion condition while this is reversed in the discussion condition where the smallest group had the highest rating. This picture is somewhat blurred by the fact that the medium-sized groups (4 persons) had the largest variability so there is no clear relation between group size and these variables.

For the remaining questionnaire items (which we cannot report here fully due to space limitations) there is a tendency in favour of the discussion version both in the items related to usability and acceptance of the system as well as concerning the fit of the recommendations and the ease with which a matching hotel could be found.

The time needed to come to a decision differed significantly between the introductory task and open task (13,500 vs. 26,333, p = 0.05). Results concerning negotiation behavior are listed in Table 2: both individual changes and number of petitions increase with group size. In relation with Table 1, it may be concluded that users in small groups are generally more satisfied because they were able to select more preferences for themselves and made less changes in their individual lists (keeping their initial wishes).

Discussion: The results of this study can only give a first indication of how well the proposed approach works in comparison to other techniques and in different group

Table 2. Objective results (lower and upper bounds at 95 % confidence interval).

	3 Participants			4 Participants			5 Participants		
	m	LB	UB	m	LB	UB	m	LB	UB
Time	21	31	10,9	17,2	7,23	27,2	21,5	11,4	31,5
Pref. Sel./Part.	3	2,37	3,62	2,31	1,76	2,85	2,80	2,31	3,28
Ind. Changes/Part.	7,33	1,79	12,8	10,1	4,64	15,7	13,1	7,61	18,6
Petitions/Part.	0,66	0	1,65	0,68	0	1,54	1,95	1,18	2,71

contexts. We can see significant advantages for our approach of including discussion and negotiation features in a group recommender in some relevant items, as well as a tendency in favour of the system in the majority of other items. However, it appears that the system may be more useful in small groups. This may be due to several factors: first, as larger groups require more communication and negotiation to obtain an acceptable end results, this may increase the complexity of the task and the interaction effort. This may be true for other group decision making systems as well but will require further research. A second factor may be artificially created by the experimental method used. Since users were instructed to play the roles described in their respective role cards, the diversity of preferences increased with group size, possibly making it more difficult to make sense of the diverse standpoints and to lead the negotiation towards a joint group decision. This may not be the case in typical real world settings where group members' viewpoints may be more homogenous due to the prior history of the group. Also, the role card method can only be taken as an approximation of a real situation. In any case, the observed tendencies raise interesting general questions concerning test scenarios for evaluating group recommender systems.

6 Conclusions and Outlook

We have presented a novel approach to group recommending that provides more interactive control over the recommendation process than typical group recommenders and that does not require the prior availability of the group members' preference profiles, taking into consideration cold-start situations and potential privacy concerns. Most importantly, the method provides discussion and negotiation support in a collaborative preference elicitation and negotiation process. Individual preferences are aggregated in a group preference profile which is immediately updated when users change preferred features or their relevance level. Also, the resulting recommendations are continuously recalculated when group preferences change, and are always visible to the whole group. Since producing recommendations constitutes just an intermediate step in the group decision process, we also support group interaction in the final decision steps where the group needs to find consensus about the item finally selected.

The proposed technique provides much higher flexibility and responsiveness to situational needs than the fixed strategies typically used in group recommenders. While this research has focused on specifying preferences in an ad hoc fashion, the method can easily be extended by storing and re-using user profiles, thus reducing interaction effort to simply adapting an existing profile. Since the preferences of other users and resulting group preferences as well as the recommendations that match this profile are always visible, participants' awareness of individual and group views and of the effects of their preference settings is increased.

Based on these concepts, we developed the prototype hotel recommender Hootle and tested it in a user study. The results indicate a higher overall satisfaction with the system as well as a higher perceived recommendation quality when compared against a system version where no discussion was possible. However, we also saw an indication of an interaction effect between group size and the two system versions which suggests that the negotiation-based approach may be more suitable for smaller groups. Whether this effect is due to the increased communication effort in larger groups, or may be dependent on the experimental scenarios used in the study is still an open question.

In future work, we aim at investigating the effects of group size more deeply and at optimizing the system to better scale for larger groups. A further work item is to consider alternative aggregation functions that may perform better than the Borda Count variant currently used. Finally, we aim at further improving the user experience with respect to the discussion and decision making features implemented. Also, more extensive empirical studies are planned, addressing also domains other than hotel selection.

References

1. Ardissono, L., Goy, A., Petrone, G., Segnan, M.: A multi-agent infrastructure for developing personalized web-based systems. ACM Trans. Internet Technol. (TOIT) **5**(1), 47–69 (2005)
2. Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. Appl. Artif. Intell. **17**(8–9), 687–714 (2003)
3. Bekkerman, P., Kraus, S., Ricci, F.: Applying cooperative negotiation methodology to group recommendation problem. In: Proceedings of Workshop on Recommender Systems in 17th European Conference on Artificial Intelligence (ECAI 2006), pp. 72–75. Citeseer (2006)
4. Beckmann, C., Gross, T.: Towards a group recommender process model for ad-hoc groups and on-demand recommendations. In: Proceedings of the 16th ACM International Conference on Supporting Group Work, pp. 329–330. ACM (2010)
5. Boratto, L., Carta, S.: State-of-the-art in group recommendation and new approaches for automatic identification of groups. In: Soro, Alessandro, Vargiu, Eloisa, Armano, Giuliano, Paddeu, Gavino (eds.) Information Retrieval and Mining in Distributed Environments. SCI, vol. 324, pp. 1–20. Springer, Heidelberg (2010)
6. Brooke, J.: SUS-A quick and dirty usability scale. Usability Eval. Ind. **189**, 194 (1996)
7. Hartnett, T.: Consensus-Oriented Decision-Making: the CODM Model for Facilitating Groups to Widespread Agreement. New Society Publishers, Gabriola Island (2011)
8. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 5–53 (2004)
9. Hill, W., Stead, L., Rosenstein, M., Furnas, G.: Recommending and evaluating choices in a virtual community of use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 194–201 (1995)
10. Jameson, A.: More than the sum of its members: challenges for group recommender systems. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 48–54. ACM (2004)
11. Jameson, A., Baldes, A., Kleinbauer, T.: Two methods for enhancing mutual awareness in a group recommender system. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 447–449. ACM (2004)
12. Jameson, A., Smyth, B.: Recommendation to groups. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 596–627. Springer, Heidelberg (2007)
13. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. User Model. User-Adap. Inter. **22**(4–5), 441–504 (2012)
14. Kompan, M., Bielikova, M.: Group recommendations: survey and perspectives. Comput. Inform. **33**(2), 446–476 (2014)

15. Lieberman, H., Van Dyke, N., Vivacqua, A.: Let's browse: a collaborative browsing agent. Knowl.-Based Syst. **12**(8), 427–431 (1999)
16. Liu, X., Tian, Y., Ye, M., Lee, W.: Exploring personal impact for group recommendation. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 674–683. ACM (2012)
17. Loepp, B., Hussein, T., Ziegler, J.: Choice-based preference elicitation for collaborative filtering recommender systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2014), pp. 3085–3094. ACM, New York (2014)
18. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers, pp. 93–141. Personalized Digital Television. Springer, The Netherlands (2004)
19. McCarthy, J.F., Anagnost, T.D.: MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, pp. 363–372. ACM (1998)
20. McCarthy, K., McGinty, L., Smyth, B.: Case-based group recommendation: compromising for success. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 299–313. Springer, Heidelberg (2007)
21. McCarthy, K., McGinty, L., Smyth, B., Salamó, M.: The needs of the many: a case-based group recommender system. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) ECCBR 2006. LNCS (LNAI), vol. 4106, pp. 196–210. Springer, Heidelberg (2006)
22. McCarthy, K., McGinty, L., Smyth, B., Salamo, M.: Social interaction in the CATS group recommender. In: Workshop on the Social Navigation and Community Based Adaptation Technologies (2006)
23. McCarthy, K., Salamo, M., Coyle, L., McGinty, L., Smyth, B., Nixon, P.: CATS: A synchronous approach to collaborative group recommendation. In: FLAIRS Conference, vol. 2006, pp. 86–91 (2006)
24. McGrath, J.E., Berdahl, J.L.: Groups, technology, and time. In: Scott Tindale, R., et al. (eds.) Theory and Research on Small Groups, pp. 205–228. Springer, US (2002)
25. Nunamaker Jr., J.F., Briggs, R.O., Mittleman, D.D., Vogel, D.R., Balthazard, P.A.: Lessons from a dozen years of group support systems research: A discussion of lab and field findings. J. Manage. Inf. Syst. **13**, 163–207 (1996)
26. O'connor, M., Cosley, D., Konstan, J.A., Riedl, J.: Polylens: a recommender system for groups of users. In: Prinz, W., et al. (eds.) ECSCW 2001, pp. 199–218. Springer, The Netherlands (2001)
27. Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.P., Jonker, C.M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. User Model. User-Adap. Inter. **22**(4–5), 357–397 (2012)
28. Pu, P., Chen, L.: User-involved preference elicitation for product search and recommender systems. AI Mag. **29**(4), 93 (2009)
29. Pu, P., Chen, L., Hu, R..: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 157–164. ACM (2011)
30. Ricci, F., Rokach, L., Shapira, B.: Introduction to Recommender Systems Handbook. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 1–35. Springer, US (2010)
31. Saaty, T.L.: Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process, vol. 6. RWS Publications, Pittsburgh (2000)
32. Walther, J.B.: Computer-mediated communication impersonal, interpersonal, and hyper personal interaction. Commun. Res. **23**(1), 3–43 (1996)

Paper II. Hootle+: A Group Recommender System Supporting Preference Negotiation

Reproduced with permission from Springer Nature: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Hootle+: A Group Recommender System Supporting Preference Negotiation*. In *Collaboration and Technology*, Lecture Notes in Computer Science, pp. 151–166.

https://doi.org/10.1007/978-3-319-44799-5_12

Hootle+: A Group Recommender System Supporting Preference Negotiation

Jesús Omar Álvarez Márquez(✉) and Jürgen Ziegler

University of Duisburg-Essen, Duisburg, Germany
{jesus.alvarez-marquez,juergen.ziegler}@uni-due.de

Abstract. This paper presents an approach to group recommender systems that focuses its attention on the group's social interaction during the formulation, discussion and negotiation of the features the item to be jointly selected should possess. The system supports a collaborative preference elicitation and negotiation process where desired item features can be defined individually, but group consensus is needed for them to become active in the item filtering process. Users can provide feedback on other members' preferences and change their significance, bringing up new recommendations each time individual settings are modified. The last stage in the decision process is also supported, when users collectively select the final item from the recommendation set. We developed the prototype hotel recommender Hootle+ and evaluated it in a user study involving groups of different size. The results indicate a good overall satisfaction, which increases with group size. However, the success ratio for bigger groups is lower than for small groups, raising questions for follow-up research.

Keywords: Group recommender system · Group preference elicitation · Negotiation · Decision-making

1 Introduction

Over the recent years, recommender systems (RS) have become an important and widely used technology that can help users in selecting items from large sets of choices, for example, in online shops or media portals [32]. RS are usually aimed at supporting individual users in their search and decision-making, which is appropriate in many cases where an item (such as a news article) is typically only utilized by a single user. Already early on, RS research recognized that there are also situations where groups of people utilize a product or service together, for example, when jointly going to a restaurant or the movies. *Polylens* was the first system that supported group decisions by providing recommendations based on the users' preferences [28]. A number of group recommender systems (GRS) have been developed since [6, 20] but there is still limited research in this area and the question of how to optimally support a group decision process based on recommendations is still open in several aspects. Usually, GRS extract the information they need from existing user profiles, subsequently using one out of two approaches for calculating the recommendations: either they aggregate the user profiles to create a single group profile (model aggregation) before generating group recommendations, or the recommendations are individually calculated for each

T. Yuizono et al. (Eds.): CRIWG 2016, LNCS 9848, pp. 151–166, 2016.
DOI: 10.1007/978-3-319-44799-5_12

user profile and then aggregated, using a variety of different strategies (recommendation aggregation). These approaches fail, however, when user preference data is not available, either for single users or for the whole group, which is the case in cold start situations. This obstacle is especially problematic for ad-hoc groups who gather spontaneously or when user data are distributed over different unconnected systems. A further issue is the situational variability of user preferences, which may amplify the inherent heterogeneity of preferences due to different responses of group members to the situational context. These issues ask for methods that can elicit group preferences on the fly and that can aggregate individual preferences in a manner that best suits the individual users as well as the group as a whole. In addition, other processes occurring in group interaction, such as developing or refining one's own preferences and requirements based on the group discussion, or negotiating with others about the desired features of an item, have so far been under-explored in GRS research.

In this paper, we present an approach to GRS that is based on the intersection of conventional recommender techniques and decision-making support for groups. In a precursor development [1], we obtained promising results but also uncovered some issues that leave space for improvement, which motivated this follow-up research and the development of a revised method and prototype. From the previous development, we kept the underlying basic idea of allowing user to collaboratively create and discuss a preference model (thus addressing collaborative preference elicitation [30]), from which recommendations are generated. Although the old system let participants generate their individual preference model by creating public lists of features ordered by importance that were subsequently aggregated into the group's model, a user study taught us that the information tended to be too complex for unexperienced users, and that it was hard for participants to keep track of the changes, an issue that became more noticeable for larger groups. With these concerns in mind, we reshaped the group interaction process in a way that users do not only change and discuss their individual preference model, but are also able to manipulate the group's preference model directly. In this process, group interaction can happen at two (tightly intertwined) stages: (1) users can online discuss and negotiate preferences proposed and accepted, and (2) they can discuss and rate items taken from the recommendation set to arrive at a final consensus decision.

A major goal in this development was to avoid unfair situations in which some users might not be satisfied with the items proposed by the system. Instead of applying a fixed strategy, as is the case in most GRS, we based our work on the assumption that computer-mediated discussion groups have more equal member participation [35]. Each user can individually specify the features the jointly selected item should possess and propose them to the group. The group decides through public voting which attributes will be accepted and rate their significance, using an explicit preference elicitation approach [29]. Features that are accepted become part of the group preference model, which is used to determine an initial set of recommendations. By group discussion, members may then be able to convince other users to modify their preferences that were included into the group model. Recommendations are continuously calculated and updated after each change, thus allowing users to see the effect of their actions immediately. Different mechanisms are provided for discussing and reaching an agreement, both for the creation of a group preference model and for the final item selection.

In the following, we first survey related research before presenting the conceptual aspects of our approach (Sects. 2 and 3). We then describe the implementation of the prototype Hootle+ and its user interface design in Sect. 4. We report on a user study performed with groups of different sizes in Sect. 5 and conclude by summarizing our work and outlining future work in Sect. 6.

2 Related Work

While the field of recommending items for single users has already received a great deal of attention in recent research, GRS are, in comparison, a still less deeply investigated area. However, various GRS have been developed over the recent years, starting from early systems such as *MusicFX* [21], a group music recommender, that use different approaches for generating recommendations [6, 14]. However, there are still many open research questions concerning, for example, the best approach to aggregating individual preferences, techniques for responding to the situational needs of the group, or supporting the social interaction processes in the group for converging on a joint decision.

To structure the wide range of different aspects involved in group recommending, [16] suggest a design space comprising the dimensions preference input, process characteristics, group characteristics, and output. In the process dimension, an important aspect is how individual, possibly conflicting preferences can be merged to obtain recommendations that best fit the group as a whole. Apart from a few exceptions, group recommenders commonly use one of two schemas for gathering and representing users' preferences [14], already mentioned during the introduction. The first one, prediction aggregation, assumes that for each item, it is possible to predict a user's satisfaction, given the user's profile; then, making use of some specific aggregation strategy, items are sorted by the group's overall satisfaction. In [11] a video recommender that uses this strategy is described; also, Polylens [28], a system that suggests movies to small groups of people with similar interests, based on the personal five-star scale ratings from Movielens [10] uses this method.

The second most used strategy, model aggregation, utilizes single user profiles for generating a group preference model, which is then employed to generate matching recommendations. There exists a high number of methods used for creating the group's model: in *Let's Browse* [17] the group preference model can be seen as an aggregation of individual preference models; in *Intrigue* [2, 3] (which recommends sightseeing destinations for heterogeneous groups of tourists) the group preference model is constructed by aggregating preference models of homogeneous subgroups within the main group; *MusicFX* [21] chooses background music in a fitness center to accommodate members' preferences, also by merging their individual models; *AGReMo* [5] recommends movies to watch in cinemas close to a location for ad-hoc groups of users, creating the group's preference model not only by individual model aggregation but also taking into account specific group variables (e.g. time, weight of each member's vote). Furthermore, the *Travel Decision Forum* [12, 13] creates a group preference model that can be discussed and modified by the members themselves, aiming to non-collocated groups who are not able to meet face to face, allowing asynchronous communication.

Regardless of whether the aggregation is made before or after generating recommendations, an aggregation method that is appropriate for the specific group characteristics needs to be chosen. There are a number of voting strategies, empirically evaluated in [20], that have been used in actual GRS. One of the most typically chosen is the average strategy, where the group's score for an item is the average rating over all individuals (e.g., it is used by *Intrigue* and *Travel Decision Forum*); on the other side, the least misery strategy scores items depending on the minimal rating it has among group members (*Polylens, AGReMo*); placed somewhere in between, the average without misery strategy consists in rating items using an average function, but discarding those where the user score is under a threshold (*MusicFX, CATS* [22–25]); as a final example of most used aggregation methods, the median strategy uses the middle value of the group members' ratings (*Travel Decision Forum*).

On another dimension, the question of preference elicitation has to be solved, which is concerned with how the user-specific preference information needed to generate recommendations is obtained. One approach is to let users rate a number of items in advance and to derive preferences from this set of ratings. *AGReMo*, for instance, requires group members to create their own model of individual preferences before the group meeting takes place by rating movies that they already saw. In *Travel Decision Forum,* each participant starts with an empty preference form that has to be filled with the desired options, so group members define new preferences for each session. A more interactive approach, although for single user systems, is described in [19], which requires users to repeatedly choose between sets of sample items that are selected based on latent factors of a rating matrix. The techniques mentioned also address the cold-start problem when no user profile is available up-front but initially require some effort on the part of the user to develop a sufficiently detailed profile.

However, most preference elicitation techniques do not considerate group interaction. As pointed out in [18], to obtain adequate group recommendations it is not only necessary to model users' individual preferences, but also to understand how a decision among group members is reached. While research on group decision-making [33] is concerned with collaboratively making choices, focusing on the social process and the outcome, these aspects have mostly not been addressed in the development of GRS. Group decision making involves a variety of aspects, such as the discussion and evaluation of others' ideas, conflict resolution and evaluating the different options that have been elaborated. Also interesting for our research is the concept of consensus decision-making [9], which seeks for an acceptable resolution for the whole group. Within this context, Group Decision Support Systems (GDSS) have emerged, that aim at supporting the various aspects of decision-making [26, 27]. Recent examples of GDSS are *Choicla* [34] (domain-independent decision-making tool) or the popular *Doodle* [8] (event scheduling). Only few GRS attempt to include aspects of group decision theory, for instance, by introducing automated negotiation agents that simulate discussions between members to generate group recommendations [4]. However, supporting the entire preference elicitation and negotiation process that may occur when users take recommender-supported decisions is, to our knowledge, not realized by current GRS.

Taking into account the social factor that is involved in group recommendation, one needs to contemplate the question whether a user would be willing to change personal preferences in favor of the group's desires, bringing up the importance of group negotiation. In the Travel Decision Forum again, users are able to explore other members' preferences, with the possibility to copy them or propose modifications. The Collaborative Advisory Travel System (CATS) focuses on collocated groups of persons gathered around a multi-touch table. Recommendations are made by collecting critiques (users' feedbacks respecting recommended destinations) that can be discussed face to face, since the system gives visual support to enhance awareness of each other's preferences. The main difference between CATS and the system we propose is that the former is focused in critiquing items once they have been recommended, whereas the latter allows negotiation already in the preference elicitation stage.

3 Preference Elicitation and Negotiation Method

The approach here described is built on the idea of letting users remotely collaborate to create the set of preferences that will conform the group's preference model. As a result, users do not discuss only about recommendations, but also about which attributes should be examined by the system when exploring the items to recommend. For doing so, preferences are evaluated in a process that involves interaction among group members almost since its very first stage until the last one. The result will be a very well narrowed set of preferences and a collection of recommended items matching the group's overall wishes. The overall process is carried out as follows:

1. Each participant can individually select the features that the recommended items should contain by placing them in a private area.
2. Once a feature is selected, the user may propose it to the rest of the group, together with the importance this user thinks that this feature deserves.
3. By proposing a feature, it becomes visible to the whole group, which will decide whether to accept it as a filter or not using the voting system provided.
4. If the feature is accepted, it becomes an active filter and influences the recommendations depending on its significance. A feature's significance is calculated by aggregating the importance level that each user has given to it. Significance is adjustable at any moment, bringing up new recommendations after any change.
5. Finally, a user is able to highlight specific recommended items and to state an opinion (via voting/discussing) about the ones that have been selected by the rest of users. More features can be proposed, accepted and rated continually, so the recommendations are narrowed until the group finds an item that satisfies their needs.

The proposition pool and the possibility to specify the filters' importance individually, having immediate feedback in the group model and the recommendations increases participants' awareness of others' preferences and the effects their own preferences have on the group results. The approach also entails aspects of critique-based recommenders since users can criticize or accept proposed features or recommended items. In contrast to fully automated recommender system, users have a higher level of control over the process and can easily adapt it to their current situational needs and context.

4 Description of the System

Following the aforementioned guidelines, a new, completely redesigned version of the Hootle GRS described in [1] was implemented. The prototype makes use of content-based techniques and is applicable to many different domains, provided properties of the items to be recommended are available. For demonstration purposes, we chose hotel selection for group travel as application area and used an Expedia dataset consisting of 151.000 hotel entries with descriptive information.

The different areas in which the interface is divided are shown in Fig. 1.

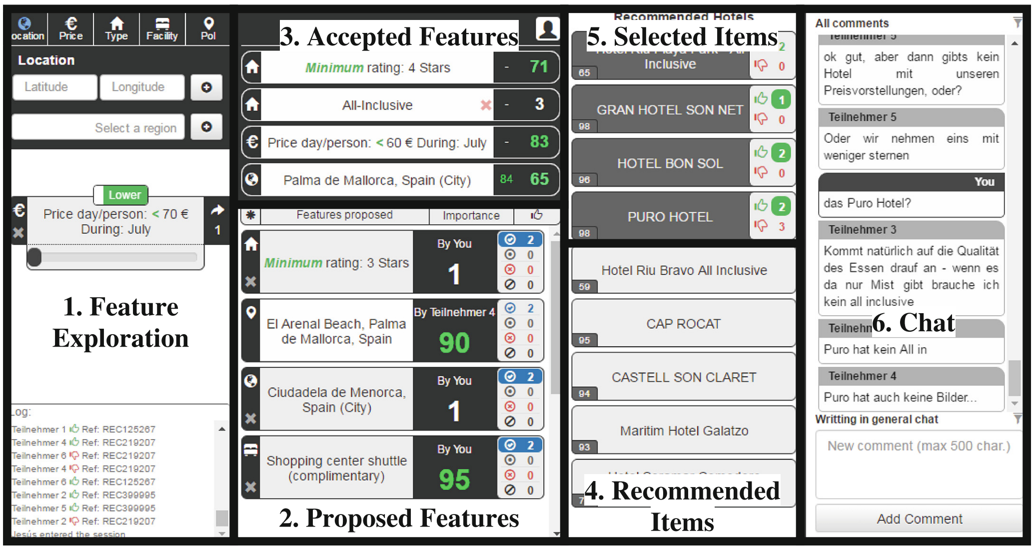


Fig. 1. Areas of the interface.

1. **Feature exploration.** Area for exploring item features by using a set of given filters (e.g. location, facilities or nearby points of interest). It is also possible to provide an importance level and to specify if the attribute is negative or positive.
2. **Proposed features.** Proposed features are shown into this area, which is shared by all participants. Voting is enabled for each proposed attribute, which can be accepted as a group filter, rejected or vetoed, depending on the results.
3. **Accepted features.** This area contains the attributes that have been approved (or vetoed) by the group. Together with their specific significance level (individually set by group members), these attributes conform the group's preference model.
4. **Recommended items.** The system calculates and displays recommendations into this area. The list is constantly updated in real-time when some group filter is added/removed or its significance changes.
5. **Selected items.** Recommended items selected by users are placed here, so other participants can see them as well.
6. **Chat.** Chat to discuss arbitrary questions that come up during the decision process. Specific discussion threads for attributes and items are provided too.

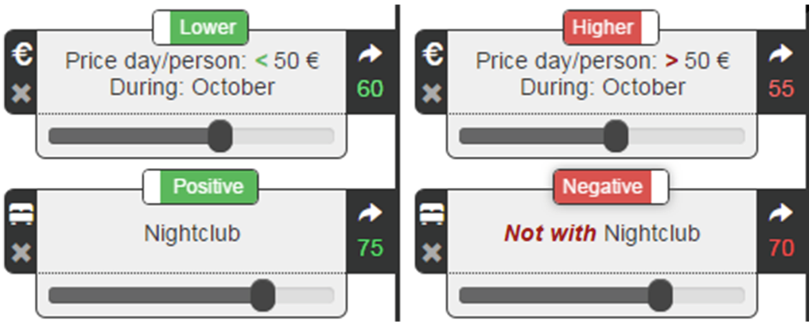


Fig. 2. Definition of two positive attributes and their negative counterparts. Importance level is specified by the slider under the attribute and displayed as a value at the right side.

4.1 Collaborative Preference Elicitation

Selecting New Attributes. This is the first step in a process that might be repeated several times. Users create new attributes by searching them through the filters located into the "feature exploration" area. Creating an attribute to propose it as a group filter (which means being part of the group's preference model) consists in selecting one of the attributes provided by the system and adding the value, type and importance attached to it (Fig. 2). Possible attribute types are *positive* (the attribute should be part of the recommendations features) and *negative* (where the opposite is preferred). For the price related attributes, "negative" and "positive" types are changed for "higher" and "lower" types. The importance level is a number between 1 and 100 that determines how relevant is the attribute in question for its creator.

Proposing an Attribute. Action that means moving a feature into the "proposed features" section of the interface, where attributes become visible for the whole group. When an attribute enters this phase, voting is enabled. Votes are not anonymous, opening the door to discussion and negotiation regarding the acceptance or rejection of the proposed features. Group members have four different choices to vote for (Fig. 3):

- To accept an attribute (blue check mark). The user acknowledges the attribute and agrees in creating a group filter from it.
- To stay neutral (grey dot). The user doesn't care about the attribute, but doesn't have any reason for not including it if others wish to do so either.

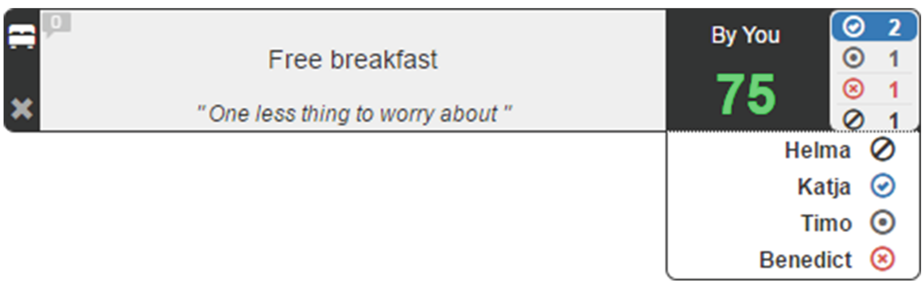


Fig. 3. Proposed attribute showing its importance level and voting results so far.

- To remove an attribute (red cross). A user manifests willingness to remove an attribute, although proposed attributes can only be removed by their creators.
- To veto an attribute (black slash). Vetoing an attribute prevents the system from using it as an active filter, even when a majority of group members has accepted it. An attribute vetoed by the whole groups becomes a veto filter and every single item containing it will be removed from the calculated recommendations.

Creating the Group Preference Model. Attributes that make it through the voting process are moved to the "accepted features" area. Features inside this area, together with their significance, conform the group preference model. Significance of an attribute is calculated using a predefined aggregation function over the importance level that each user has given to the attribute in question. Individually assigned importance levels are public knowledge among the group (Fig. 4).

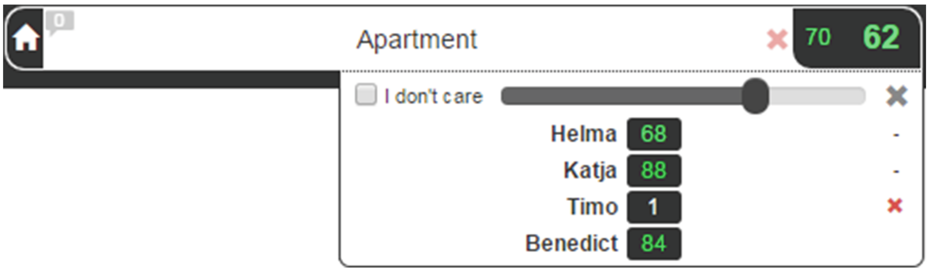


Fig. 4. Accepted feature. Individual importance levels and group significance are displayed.

An attribute that has been already accepted can be removed if the majority of the group want to do so. A removed group filter is returned to the proposed features area.

4.2 Generating Recommendations

The system takes the given preference model and explores the DB using a content-based filtering method (Fig. 5). In content-based filtering, items are described by a set of attributes, which are compared against the preference model of a user (in our case, the collaboratively created group model). Because the preference model is created from scratch in each new session, the system is applicable in cold-start situations where no user profile exists yet. Items in the DB are scored depending on how many positive attributes they contain and their significance (items with negative attributes will receive negative scores, while items containing vetoed features are removed). Once the items have been rated, the system extracts those with the highest scoring.

Every time that the group's preference model changes, new recommendations are obtained, enabling real time feedback. It could happen that none of the collected items completely fulfills the group model. In the case that only the top rated items were selected, it would be possible that for some of the attributes inside the group preference model not a single matching recommendation were provided. Because the system's

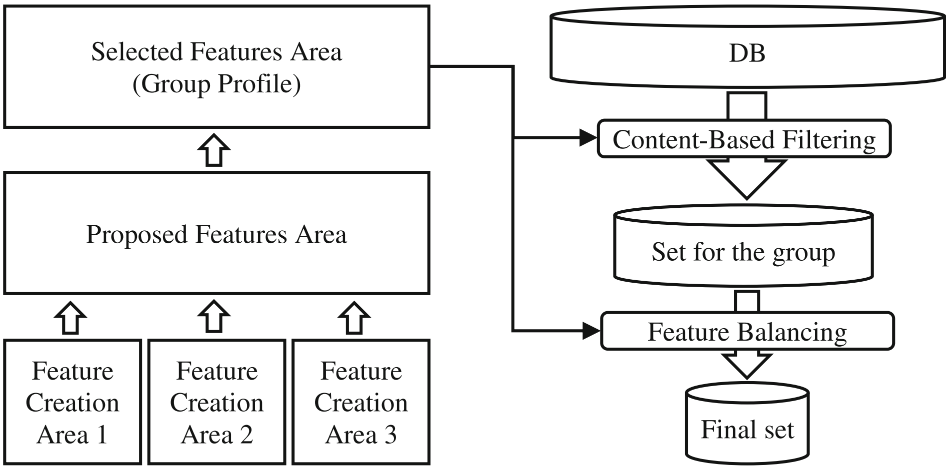


Fig. 5. Scheme of the filtering process.

raison d'etre is to serve as a tool for discussion and consensus finding within the context of GRS, it makes sense to try to return a well-balanced set of recommendations, allowing these who have chosen less popular attributes to be an active part of the negotiation process. Thus, a further step is done before sending the found recommendations to the session's participants, attempting to collect a set of items where there is at least one fitting item per attribute in the preference model.

As firstly said, the system does not require of any previously stored user profiles, something of a great usefulness when dealing with a group situation for the reasons mentioned earlier. Nevertheless, there is still plenty of room for expanding the method with more complex and longer-term user profiles, built upon the user's past choices. The interaction effort needed to specify the desired features could be lightened by starting the session with some auto-generated proposed features or letting the system elaborate a preset group preference model. Increasing the precision of the recommendations could also be a possibility by using attributes that participants have not defined for the current session, but knowing that they were selected in the past.

4.3 Negotiation

Many of the preferences stated by a user depend to a great extent on the situational context of the group and the social interaction that takes place within it. Opinions can be influenced by others through negotiation, making possible the reconciliation of adverse points of views. Thus, group decision making is an important part of the process and group agreement may not be found without an appropriate set of tools supporting discussion, negotiation and consensus finding.

Communication. Being able to talk, explaining the own decisions and questioning the reasons of others are fundamental actions in group decision making; therefore, written communication (for non-collocated groups) is of great importance. It is supported via chat and enhanced by other mechanisms, detailed in the following paragraphs.

- *Chat.* A general chat is provided where users can discuss questions that involve the whole process. Besides it, specific discussion threads for each attribute and recommendation are available too, keeping comments organized.
- *Significance.* A visual mechanism for expressing an opinion in relation with a particular feature. Each attribute has a slider that allows the users to individually define how important a feature is for them (within a scale from 1 to 100). This action, besides helping the system to generate the recommendations, provides to the rest of users a quick view of who likes and who dislikes an attribute.
- *Voting.* Users can express their consent to accept/remove/veto an attribute by voting. Votes are not anonymous, which means that a user knows at any moment what the others think about the feature at issue, giving them the chance to convince the other members and negotiate the outcome of the polling. Much the same happens with the recommended items, where users are able to vote recommendations up and down.

Conflict Resolution. A conflict appears when two or more participants want features that contradict each other. This situation is reflected by the system when the same attribute is proposed twice, once positively and once negatively, or when two incompatible (but different) attributes are added. For the first case, when one of the attributes is accepted by the group as filter, the other one is removed and no further discussion is needed. However, the second circumstance is a little bit trickier, because in many cases the system is not able to notice the contradiction by itself and the task of dealing with them relies on the users. As an attempt to support the participants visually, they have access to information about each recommended item. Those entries in the group preference model that are not fulfilled by an item are highlighted in red color, so a user can easily tell apart the conflictive attributes and try to change the opinion of the members who added them, with the expectation of removing them or lowering their significance.

4.4 Towards the Right Decision

Finding a recommendation that matches the group whishes may require several tries. Usually, it will be necessary to move through the different stages of the process in a cyclic and iterative fashion, modifying the group preference model and exploring the new recommended items once again. When negotiation and discussion are the driving force of this changes, with each new iteration the group should get closer to a solution, optimizing the group filters and narrowing down the recommendations.

Nevertheless, even when the process is carried out properly, the criteria for selecting the "right item" may differ from one scenario to another: in some cases, it could be the one that has been accepted by the majority; in others, it could be unacceptable to choose an item that has been rejected by only one member of the group. While a fixed group recommendation strategy might be used, we believe that the system cannot generally resolve such decision problems. Our approach provides tools for preference specification, discussion and acceptance measuring, but it is not possible to talk about the one right solution when dealing with group decision making in a real life situation. Ultimately, it is up to the users to decide whether a recommendation fits their needs or not and to make the final choice.

5 Evaluation

To evaluate our approach, we performed a user study with several groups comprising either three or six users, which is the range of group sizes we expect to occur in real applications. In a user study with the previous system version, we noticed an interesting correlation effect between group size and satisfaction, but had groups of three, four and five members, which may have limited the reliability of the results due to the limited range. We thus decided to slightly increase the range and focus on the extreme values. We also set up a group who used a limited version of the system with discussion facilities disabled as control, but for practical reasons could only set up one group of each size, leading to inconclusive results that are not further considered here. The main objectives of this study were to determine the usability of the approach and the quality of the resulting recommendations, as well as, more specifically, to analyze the impact of the cooperative preference elicitation and negotiation tools developed.

5.1 Setting and Experimental Tasks

We made use of a hotel database provided by Expedia with 151,000 entries. For each hotel, a full description and a set of attributes, including property and room amenities (within 360 possibilities), locations (258,426) and points of interest nearby (94,512) were available. We prepared two task scenarios with different levels of complexity:

- In an 'introductory' task, the group was instructed to select a hotel knowing beforehand some common, desired attributes, as well as the location of the hotel. This task also served as a training session to allow participants to explore the functions and possibilities the system supplies. The following scenario was presented –*"Your group will be participating at a conference in Berlin. As the conference always provides lunch and dinner, you just need to find a hotel including breakfast. Your conference will take place near the Brandenburg Gate."*
- In the 'open' task, which was always performed after the introductory task, only un-specific instructions were given to the group. The scenario used for this task was – *"It is summertime. You and your friends really need to get out of the daily routine. Discuss where to stay."*

To prevent participants from complying too quickly with the wishes of other users, we artificially induced different backgrounds and objectives for each group member. For this purpose, we created a set of role cards for the second task that were randomly distributed among group's members, with the intent of generating conflicts and discussion. A problem detected in the precursor study was that the roles used were so different one from each other that in many cases they created an artificial situation that is not commonly found in real life, where groups that plan to travel together tend to share similar preferences. Thus, for this occasion the roles were simplified and created with shared characteristics:

1. You love shopping and you are interested in cultural things.
2. You are interested in cultural things and clubbing.
3. You love partying every night. During the day, shopping keeps you awake.
4. You like to spend your time on the beach. When that is not possible, hiking fits well.
5. You prefer to hike the whole day and do sport related activities.
6. You are a sport addict and you love the beach.

5.2 Method

39 people (22 females, 17 males, average age of 22.63, σ 3.65) took part in the study, distributed in 5 groups of 3 participants and 4 groups of 6. Since the system is web-based, all users were provided with a normal desktop computer with a display screen of 21" and running the same browser. They sat in a large lab room but were separated from each other and instructed to communicate only via the means provided by the system.

Each group first received a brief introduction to the system and was asked to work on the two decision tasks, always in the order introductory task – open task. Before beginning the second task, they all received randomly one of the role cards. A task was considered complete when the group found consensus (i.e. agreed on a hotel) or the time ran out (25 min maximum per task).

After completing both tasks, participants were asked to fill in a questionnaire regarding aspects such as the quality of the recommendations or the ease-of-use of the system, using a 1–5 scale. It comprised the SUS items [7] as well as items from two recommender-specific assessment instruments (User experience of RS [15] and ResQue [31]). The recommender-specific items measure the constructs *user-perceived recommendation quality, perceived system effectiveness, interface adequacy, and ease of use*.

5.3 Results and Discussion

Not all groups were able to find a solution, reaching the time limit for the tasks. For the 3 person groups, agreement was always achieved in contrast to the 6 person groups, where only a 25 % of the tasks were completed with consensus regarding the item to select. An average success rate over all sessions of 66 % was reached. Despite the low success ratio for the bigger groups, the percentage of agreement among users (participants who selected the same hotel) was 77 %, as shown in the objective data listed in Table 1. Time needed per task was higher for the 6 people groups, as well as the amount of individual preference changes made per user (importance level, vote selection), but the number of comments written per user in the bigger groups was lower than in 3 people groups. This could mean that participants in bigger groups made a more extensive use of the graphical interface for showing their wishes and opinions to the rest of the group, because relying only in chat communication for transmitting ideas is usually more complicated the more people are writing at the same time. Despite these differences, both group types elaborated preference models with similar sizes.

Table 1. Objective results. Lower (LB) and upper (UB) bounds at 95 % confidence interval.

	3 people groups			6 people groups			Avg.
	m	LB	UB	m	LB	UB	m
Time per task (minutes)	13,60	10.18	17.01	17,63	13.8	21.43	15,61
Preference Model Size	6,10	3.85	8.34	6,38	3.87	8.88	6,23
Changes per user	12,33	6.123	18.54	14,56	11.09	18.03	14,35
Comments per user	7,16	2.42	11.90	6,41	3.77	9.06	6,92
Solution found	100 %	–	–	25 %	–	–	62.5 %
Agreement among users	100 %	–	–	77 %	–	–	88 %

In relation to the usability of the system, it received a SUS score of 65, placing the prototype slightly under the average. An independent-samples t-test was conducted to compare the items of the questionnaire, taking group size as independent variable. While many items did not show big difference between cases (Table 2), some conclusions can be extracted from them. In general, it seems harder for bigger groups to find recommendations that match the participants' individual wishes and to agree with the rest of members, which is a logical consequence of group size increase. Interesting is the fact that the groups of 6 are in general more satisfied with the tool than the smaller groups, despite being easier for the latter to find a solution through consensus.

Discussion. The outcome of the evaluation seems to indicate that some of the issues found during our previous study have been lessened, specifically the one related with how well the system scales up with group size. Even if having bigger groups increases the complexity of the decision-making process, the results point to a greater satisfaction and sense of helpfulness when using the system. This is more noticeable when one looks to the preference model size, which is almost the same through group sizes

Table 2. Some results of the evaluation.

Group size	3		6		Avg.	
	m	σ	m	σ	m	σ
The recommended items fitted my preferences	4.00	0.50	3.83	1.16	3.88	1.02
I liked the items recommended by the system	3.78	0.83	3.79	0.88	3.79	0.86
*It was very easy to find a good solution together	3.78	1.09	2.62	1.31	2.94	1.34
The other team mates agreed my opinion	4.00	0.70	3.29	1.19	3.48	1.12
*Even with different opinions we could find a good compromise	4.44	0.73	3.46	1.06	3.73	1.06
I can make a better choice with the system	3.78	0.97	3.96	1.2	3.91	1.18
I can find a solution in less time using the system	3.56	1.33	4.04	1.08	3.91	1.15
I think the program is easy to use	3.67	0.87	3.46	1.06	3.52	1.00
I think the functions in this program are well integrated	3.56	0.88	4.00	0.72	3.88	0.78
In general, I am satisfied with the system	3.56	1.13	4.33	0.96	3.76	1.00

*Significant ($p < 0.05$)

indicating that users limited the number of preferences expressed in a well-considered manner in order to facilitate consensus finding. The low ratio of solutions found for the 6 people groups could be explained as a consequence of limiting the time to finishing a task to only 25 min, but further research may be needed in order to obtain some final conclusions. In a real world situation, where the time span for finding a solution in a non-collocated group setting could be days or even weeks, and where individual preferences may tend to be more homogenous without artificially inserting roles, a higher success ratio would be expected.

6 Conclusion and Outlook

We have presented an approach to group recommended systems, which enables collaborative preference elicitation on the fly, avoiding a cold-start situation and providing more control during the recommendation process. The system supports negotiation and discussion during the preference elicitation and item selection phases. Participants can freely define and propose features, adding them to a shared pool of attributes where the group will collaboratively select those to conform the group preference model. Once the attributes are extracted, users are able to individually assign an importance level to each one of them and the system calculates their significance to the group. Recommendations are then generated after the given group preference model and will be recalculated each time that it changes. Recommendations are shown to the group members, letting them to select and discuss about those that they like, or to redefine the group preference model to obtain new recommended items.

The technique here described provides higher flexibility and awareness than the fixed strategies typically used in group recommenders. Since preferences and matching recommendations are always visible, participants' awareness of individual and group views and of the effects of their preference settings is increased.

Based on prior work, a novel prototype version of a hotel group recommender Hootle+ was developed, following the ideas described above. The results of the user study we conducted show that the new system appears to handle bigger groups better than the previous system version which did not allow users to influence the group model directly. On the other hand, we obtained a lower success rate per session, which may be due to tighter time constraints.

A work in progress is the idea of having different privileges levels defined within a session, which could be assigned to participants so their opinions would have distinct weights when voting or calculating the significance of an attribute (e.g. expert's opinion). This feature would also allow creating personalized rules for vote counting in relation to the acceptance or rejection of a feature, conferring even more flexibility to the system. It is planned to add moderator specific functions too, enabling a user to control the session's flow and to take the final decision. In future work, we will also further improve the usability of the interface, which raised some negative comments in the study. Furthermore, a detailed empirical comparison to a suitable baseline system is planned. In addition, receiving feedback from real groups of users would be a solution to the problem inherent to the use of artificial roles during the test sessions, so we are considering an online version with a realistic use case for future research.

References

1. Márquez, J.O.A., Ziegler, J.: Preference elicitation and negotiation in a group recommender system. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 20–37. Springer, Heidelberg (2015)
2. Ardissono, L., Goy, A., Petrone, G., Segnan, M.: A multi-agent infrastructure for developing personalized web-based systems. ACM Trans. Internet Technol. (TOIT) **5**(1), 47–69 (2005)
3. Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. Appl. Artif. Intell. **17**(8-9), 687–714 (2003)
4. Bekkerman, P., Kraus, S., Ricci, F.: Applying cooperative negotiation methodology to group recommendation problem. In: Proceedings of Workshop on Recommender Systems in 17th European Conference on Artificial Intelligence (ECAI 2006), pp. 72–75. Citeseer (2006)
5. Beckmann, C., Gross, T.: Towards a group recommender process model for ad-hoc groups and on-demand recommendations. In: Proceedings of the 16th ACM International Conference on Supporting Group Work, pp. 329–330. ACM (2010)
6. Boratto, L., Carta, S.: State-of-the-art in group recommendation and new approaches for automatic identification of groups. In: Soro, A., Vargiu, E., Armano, G., Paddeu, G. (eds.) Information Retrieval and Mining in Distributed Environments. SCI, vol. 324, pp. 1–20. Springer, Heidelberg (2010)
7. Brooke, J.: SUS-A quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
8. Doodle AG. http://www.doodle.com
9. Hartnett, T.: Consensus-Oriented Decision-Making: the CODM Model for Facilitating Groups to Widespread Agreement. New society publishers, Gabriola (2011)
10. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 5–53 (2004)
11. Hill, W., Stead, L., Rosenstein, M., Furnas, G.: Recommending and evaluating choices in a virtual community of use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 194–201 (1995)
12. Jameson, A.: More than the sum of its members: challenges for group recommender systems. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 48–54. ACM (2004)
13. Jameson, A., Baldes, A., Kleinbauer, T.: Two methods for enhancing mutual awareness in a group recommender system. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 447–449. ACM (2004)
14. Jameson, A., Smyth, B.: Recommendation to groups. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 596–627. Springer, Heidelberg (2007)
15. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. User Model. User-Adap. Inter. **22**(4–5), 441–504 (2012)
16. Kompan, M., Bielikova, M.: Group recommendations: survey and perspectives. Comput. Inform. **33**(2), 446–476 (2014)
17. Lieberman, H., Van Dyke, N., Vivacqua, A.: Let's browse: a collaborative browsing agent. Knowl. Based Syst. **12**(8), 427–431 (1999)
18. Liu, X., Tian, Y., Ye, M., Lee, W.: Exploring personal impact for group recommendation. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 674–683. ACM (2012)

19. Loepp, B., Hussein, T., Ziegler, J.: Choice-based preference elicitation for collaborative filtering recommender systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2014), pp. 3085–3094. ACM, New York (2014)
20. Masthoff, J.: Group modeling: Selecting a sequence of television items to suit a group of viewers. In: Ardissono, L. (ed.) Personalized Digital Television. HCI, vol. 6, pp. 93–141. Springer, Heidelberg (2004)
21. McCarthy, J.F., Anagnost, T.D.: MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, pp. 363–372. ACM (1998)
22. McCarthy, K., McGinty, L., Smyth, B.: Case-based group recommendation: compromising for success. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 299–313. Springer, Heidelberg (2007)
23. McCarthy, K., McGinty, L., Smyth, B., Salamó, M.: The needs of the many: a case-based group recommender system. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) ECCBR 2006. LNCS (LNAI), vol. 4106, pp. 196–210. Springer, Heidelberg (2006)
24. McCarthy, K., McGinty, L., Smyth, B., Salamo, M.: Social interaction in the CATS group recommender. In: Workshop on the Social Navigation and Community Based Adaptation Technologies (2006)
25. McCarthy, K., Salamo, M., Coyle, L., McGinty, L., Smyth, B., Nixon, P.: CATS: a synchronous approach to collaborative group recommendation. In: FLAIRS Conference, pp. 86–91 (2006)
26. McGrath, J.E., Berdahl, J.L.: Groups, technology, and time. In: Tindale, R.S., Heath, L., Edwards, J., Posavac, E.J., Bryant, F.B., Suarez-Balcazar, Y., Henderson-King, E., Myers, J. (eds.) Theory and Research on Small Groups. HCI, vol. 6, pp. 205–228. Springer, New York (2002)
27. Nunamaker Jr., J.F., Briggs, R.O., Mittleman, D.D., Vogel, D.R., Balthazard, P.A.: Lessons from a dozen years of group support systems research: a discussion of lab and field findings. J. Manage. Inf. Syst. **13**, 163–207 (1996)
28. O'connor, M., Cosley, D., Konstan, J.A., Riedl, J.: PolyLens: a recommender system for groups of users. In: Prinz, W., Jarke, M., Rogers, Y., Schmidt, K., Wulf, V. (eds.) ECSCW 2001, pp. 199–218. Springer, Heidelberg (2001)
29. Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.P., Jonker, C.M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. User Model. User-Adap. Inter. **22**(4–5), 357–397 (2012)
30. Pu, P., Chen, L.: User-involved preference elicitation for product search and recommender systems. AI Mag. **29**(4), 93 (2009)
31. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 157–164. ACM (2011)
32. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 1–35. Springer, New York (2010)
33. Saaty, T.L.: Fundamentals of decision making and priority theory with the analytic hierarchy process, vol. 6. RWS Publications, Pittsburgh (2000)
34. Stettinger, M.: Choicla: towards domain-independent decision support for groups of users In: Proceedings of the 8th ACM Conference on Recommender systems, pp. 425–428 (2014)
35. Walther, J.B.: Computer-mediated communication impersonal, interpersonal, and hyper personal interaction. Commun. Res. **23**(1), 3–43 (1996)

Paper III. Negotiation and Reconciliation of Preferences in a Group Recommender System

Reproduced with permission from the Information Processing Society of Japan: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Negotiation and Reconciliation of Preferences in a Group Recommender System*. Journal of Information Processing, 26:186–200.

https://doi.org/10.2197/ipsjjip.26.186

Journal of Information Processing Vol.26 186–200 (Feb. 2018)

[DOI: 10.2197/ipsjjip.26.186]

Negotiation and Reconciliation of Preferences in a Group Recommender System

Jesús Omar Álvarez Márquez[1,a)] Jürgen Ziegler[1,b)]

Received: May 24, 2017, Accepted: November 7, 2017

Abstract: This article presents an approach to group recommender systems that focuses its attention on the group's social interaction during the formulation, discussion and negotiation of the features the item to be jointly selected should possess. Current group recommender techniques are mainly based on aggregating existing user profiles or on a profile of the group as a whole. Our method supports collaborative preference elicitation and negotiation process where desired features have to be chosen individually, but group consensus is needed for them to become active in the item filtering process. Users provide feedback on the selected preferences and change their significance, bringing up new recommendations each time individual settings are modified. The last stage in the decision process is also supported, when users collectively select the final item from the recommendation set. We explored the possible benefits of our approach through the development of three prototypes, each based on a different variant of the approach with a different emphasis on private and group-wide preference spaces. They were evaluated with user groups of different size, addressing questions regarding the effectiveness of different information sharing methods and the repercussion of group size in the recommendation process. We compare the different methods and consolidate the findings in an initial model of recommending for group.

Keywords: group recommender system, group preference elicitation, negotiation, decision-making

1. Introduction

Recommender systems (RS) are well-established tools that aim at supporting users in choosing items, such as products, movies or hotels, from large sets of alternatives [36]. RS are widely applied in applications such as online shops, news portals, or media platforms and have been shown to have strong commercial implications, e.g., by increasing the number of sales [32]. A wide range of recommender techniques have been developed, both in academia and industry, that are mostly based either on users' ratings of items (provided explicitly by the user or implicitly based on interaction behavior or purchases) which is known as collaborative filtering, or on properties of the items themselves (content-based filtering). Classical approaches to collaborative filtering apply k-nearest neighbor techniques for identifying users with a similar rating behavior and predicting the user's rating for unknown items through weighted averages of similar users' ratings. Although the basic techniques have been refined and expanded over the years, a major assumption in most of them is that users have personal preferences that are stable and do not change over time. While this assumption may be considered questionable in the case of single-user recommendations (and has been abandoned in several research works), it is even more problematic if one wants to recommend items to a group of persons. There are numerous situations where the decision to buy or use a particular product or service needs to be taken by a group of people, for example, when jointly going to a restaurant or to the movies. The complexity of arriving at a joint decision acceptable to all group members is mostly higher than in the individual case since the preferences of the group members will typically differ and may be hard to reconcile. It is indeed not obvious what the preferences of a group are and how they may be derived from the preferences of their individual members. Due to the communication and social interaction in a group that happens before taking a joint decision, the overall preferences of a group tend to be more dynamic than in the single-user case and often only emerge in the group interaction process. This aspect needs to be taken into account when designing group recommender systems, but it has not yet received sufficient attention in that specific research field.

Already early on, RS research recognized that RS may have a role in facilitating group decisions, provided they offer appropriate functions for dealing with diverse user preferences and the characteristics of group decision processes. Polylens was the first system that supported group decisions by providing recommendations based on the users' preferences [31]. A number of group recommender systems (GRS) have been developed since Refs. [7], [22] but there is still limited research in this area and the question of how to optimally support a group decision process based on recommendations is still open in several aspects. Usually, GRS extract the information they need from existing individual user profiles, subsequently using one of two approaches for calculating the recommendations: either they aggregate the user profiles to create a single group profile (model aggregation) before generating group recommendations, or the recommendations are individually calculated for each user profile and then

1 University of Duisburg-Essen, Duisburg, Germany
a) jesus.alvarez-marquez@uni-due.de
b) juergen.ziegler@uni-due.de

aggregated, using a variety of different strategies (recommendation aggregation). These approaches fail, however, when user preference data are not available, either for single users or for the whole group, which is the case in cold start situations. This obstacle is especially problematic for ad-hoc groups who gather spontaneously or when user data are distributed over different unconnected systems. In addition, the situational variability of user preferences is higher than in the single-user case, amplifying the inherent heterogeneity of preferences due to different responses of group members to the situational context. A general problem with existing approaches is that they typically only consider the interaction among group members in a late phase of the process where recommendations have already been calculated and the group needs to decide which of the recommendations to select. In real situations, however, the interaction in the group tends to begin much earlier when group members, for example, articulate their preferences, try to convince others to share them, or revise their preferences to enable the group to come to a joint conclusion. In some cases, individual preferences will only emerge during this process of social interaction. These issues ask for methods that can elicit group preferences on the fly and that can aggregate individual preferences in a manner that best suits the individual users as well as the group as a whole. In addition, other processes occurring in group interaction, such as developing or refining one's own preferences and requirements based on the group discussion, or negotiating with others about the desired features of an item, have so far been under-explored in GRS research.

In this paper, we present an approach to GRS that aims at supporting the entire process of group-decision making. Our approach provides a novel contribution by focusing particularly on the early phases of group decision making, incorporating features for preference negotiation, discussion and reconciliation. The group preference profile emerging in this process is continuously fed into a recommender system that suggests items which can then be voted on or weighted by the group. We investigated these concepts in three successive prototype developments which we evaluated in empirical user studies with groups of different size. With these developments, we aimed at answering the following research questions:

- What are effective means for supporting the formulation, the exchange and the negotiation of user preferences in a distributed GRS?
- How to structure this process with respect to private spaces for setting up one's individual preferences versus public spaces that can be seen and criticized by the whole group?
- How does group size affect the usability and the acceptance of the approach and the different techniques?

Instead of applying a fixed strategy, as is the case in most GRS, we based our work on the assumption that computer-mediated discussion groups have a more equal member participation [39]. Following this idea, our approach allows a group of users to collaboratively create and discuss a preference model (thus addressing collaborative [34] and explicit [33] preference elicitation). A first prototype was designed [1] where users were able to create their own individual lists of features ordered by importance, obtaining immediate feedback on the aggregated group's preference model and its matching recommended items. The results obtained from the consequent user evaluation were promising, suggesting that our approach effectively improves the quality of recommendations when compared against standard group recommender systems. However, these results also brought to light some issues, mostly related with the performance of the approach regarding group scalability and the complexity of the displayed information, motivating a first revision of the method and the creation of a second prototype [2]. For this prototype, the method was reshaped in a way that users do not create their preference models individually, but each one of them can specify the preferred features of the item to select and propose them to the group. The group decides through public voting which attributes will be accepted and weights their significance, building the group's preference model together. In an empirical study of the revised prototype, results with respect to group scalability showed a considerable improvement. Nevertheless, new concerns appeared as well, in this case in relation to the dichotomy between private and public areas (within the tool's workspace) and if such a distinction is beneficial at all for the recommendation process. These findings led to the development of a third prototype, Hootle Mobile, based on a revised, streamlined method where private spaces have been completely removed and preferences could be directly added to the group model.

This paper provides an aggregated and extended account of work reported in a prior publication [2], incorporating a design synopsis and empirical findings from a first version of the system [1] as well as more details on its empirical evaluation. In addition, we report for the first time on a mobile version of the system that also modifies the approach by directly expressing user preferences in the shared group space. We also present the results of an empirical evaluation of this mobile version of the system.

In the following, we first survey related research and enumerate the basics of our approach (Sections 2 and 3). In Section 4, the first version of our method is described, followed by the prototype GRS Hootle based on it and the results of its study. Section 5 presents the conceptual aspects of the revised approach, its implementation in a second prototype (Hootle+) and the pertinent results of a new evaluation. The final version of the method is reported in Section 6, together with a last implementation adapted to mobile devices (Hootle Mobile). We conclude by summarizing our work and outlining further research possibilities in Section 8.

2. Related Work

While the field of recommending items for single users has already received a great deal of attention in recent research, GRS are, in comparison, a still less deeply investigated area. However, various GRS have been developed over the recent years, starting from early systems such as MusicFX [23], a group music recommender, that uses different approaches for generating recommendations [7], [16]. However, there are still many open research questions concerning, for example, the best approach to aggregating individual preferences, techniques for responding to the situational needs of the group, or supporting the social interaction processes in the group for converging on a joint decision.

To structure the wide range of different aspects involved in group recommending, Ref. [18] suggest a design space comprising the dimensions preference input, process characteristics, group characteristics, and output. In the process dimension, an important aspect is how individual, possibly conflicting preferences can be merged to obtain recommendations that best fit the group as a whole. Apart from a few exceptions, group recommenders commonly use one of two schemas for gathering and representing users' preferences [16], already mentioned in the introduction. The first one, prediction aggregation, assumes that for each item, it is possible to predict a single user's satisfaction, given the user's profile; then, through some specific aggregation strategy, items are sorted by the group's overall satisfaction. In Ref. [13] a video recommender that uses this strategy is described; also, Polylens [31], a system that suggests movies to small groups of people with similar interests, based on the personal five-star scale ratings from Movielens [12] uses this method.

The second most used strategy, model aggregation, utilizes single user profiles for generating a group preference model, which is then employed to generate matching recommendations. There exists a large number of methods for creating the group's model: in Let's Browse [19] the group preference model can be seen as an aggregation of individual preference models; in Intrigue [3], [4] (which recommends sightseeing destinations for heterogeneous groups of tourists) the group preference model is constructed by aggregating preference models of homogeneous subgroups within the main group; MusicFX [23] chooses background music in a fitness center to accommodate members' preferences, also by merging their individual models; AGReMo [5] recommends movies to watch in cinemas close to a location for ad-hoc groups of users, creating the group's preference model not only by individual model aggregation but also taking into account specific group variables (e.g. time, weight of each member's vote). Furthermore, the Travel Decision Forum [14], [15] creates a group preference model that can be discussed and modified by the members themselves, aiming to non-collocated groups who are not able to meet face to face, allowing asynchronous communication.

Regardless of whether the aggregation is made before or after generating recommendations, an aggregation method that is appropriate for the specific group characteristics needs to be chosen. There are a number of voting strategies, empirically evaluated in Ref. [22], that have been used in actual GRS. One of the most typically chosen is the average strategy, where the group's score for an item is the average rating over all individuals (e.g., used by Intrigue and Travel Decision Forum); on the other side, the least misery strategy scores items depending on the minimal rating it has among group members (Polylens, AGReMo); placed somewhere in between, the average without misery strategy consists in rating items using an average function, but discarding those where the user score is under a threshold (MusicFX, CATS [24], [25], [26], [27]); as a final example of most used aggregation methods, the median strategy uses the middle value of the group members' ratings (Travel Decision Forum).

On another dimension, the question of preference elicitation has to be solved, which is concerned with how the user-specific preference information needed to generate recommendations is obtained. One approach is to let users rate a number of items in advance and to derive preferences from this set of ratings. AGReMo, for instance, requires group members to create their own model of individual preferences before the group meeting takes place by rating movies that they already saw. In Travel Decision Forum, each participant starts with an empty preference form that has to be filled with the desired options, so group members define new preferences for each session. A more interactive approach, although for single user systems, is described in Ref. [21], which requires users to repeatedly choose between sets of sample items that are selected based on latent factors of a rating matrix. The techniques mentioned also address the cold-start problem when no user profile is available up-front but initially require some effort on the part of the user to develop a sufficiently detailed profile.

However, most preference elicitation techniques do not considerate group interaction. As pointed out in Ref. [20], to obtain adequate group recommendations it is not only necessary to model users' individual preferences, but also to understand how a decision among group members is reached. While research on group decision-making [37] is concerned with collaboratively making choices, focusing on the social process and the outcome, these aspects have mostly not been addressed in the development of GRS. Group decision making involves a variety of aspects, such as the discussion and the evaluation of others' ideas, the conflict resolution and the assessment of the different options that have been elaborated. Also interesting for our research is the concept of consensus decision-making [11], which seeks for an acceptable resolution for the whole group. Within this context, Group Decision Support Systems (GDSS) have emerged, that aim at supporting the various aspects of decision-making [28], [30]. Recent examples of GDSS are Choicla [38] (domain-independent decision-making tool) or the popular Doodle [9] (event scheduling). Only few GRS attempt to include aspects of group decision theory, for instance, by introducing automated negotiation agents that simulate discussions between members to generate group recommendations [6]. However, supporting the entire preference elicitation and negotiation process that may occur when users take recommender-supported decisions is, to our knowledge, not realized by current GRS.

Taking into account the social factor that is involved in group recommendation, one needs to contemplate the question whether a user would be willing to change personal preferences in favour of the group's desires, bringing up the importance of group negotiation. In the Travel Decision Forum again, users are able to explore other members' preferences, with the possibility to copy them or propose modifications. The Collaborative Advisory Travel System (CATS) focuses on collocated groups of persons gathered around a multi-touch table. Recommendations are made by collecting critiques (users' feedbacks respecting recommended destinations) that can be discussed face to face, since the system gives visual support to enhance the awareness of each other's preferences. In a similar fashion, the more recent STSGroup system described in Ref. [29], assists a non-collocated group of people in collaboratively finding POIs by letting them influence the outcome of the recommendations through a critiquing-based technique that works at the item level, tracking the reactions of participants when the items are proposed in the discussion chat. The main difference between these two last systems and the system we propose is that they are focused in critiquing items once they have been recommended, whereas our approach allows negotiation already in the preference elicitation stage.

3. Concept

A major objective of our work is to support all stages of group decision processes that are facilitated by group recommender systems. In contrast to existing GRS research, we therefore put a stronger focus on the initial phases of the process where users formulate their preferences and may discuss and negotiate these with other group members. To create a group recommender system that it is consistently supported by group decision theory during all the stages of the recommending process, we built our approach over three fundamental pillars:

(1) A group of non-collocated users collaborate during the preference elicitation stage for creating a shared preference model, which will be then utilized for generating group recommendations. When obtaining the items to recommend, not only the group's preference model is taken into account, but their individual preferences too.

(2) Users can, at any moment, discuss and negotiate about which attributes should be examined by the system, molding the group's preferences through group interaction. Changes made in this fashion provide immediate feedback about their effect by updating the set of recommendations.

(3) Users can discuss about the recommended items until consensus is found, thus supporting the last part of the recommendation process too.

Based on the aforementioned concepts, a method has been developed through three different iterations with three different prototypes, each one of them used to redefine the original technique after learning from the issues found during their evaluations. Also, with the different prototypes we explored different ways of structuring the process into private preference formulation and public, group-wide visibility and negotiation of preferences.

In our first approach, we prioritize transparency through the recommending process by supporting single user preference elicitation, letting each participant specify his or her own individual preference model by selecting a number of desired attributes and ranking them by importance, being all of these models aggregated to create the group's one. The only way they have to influence the system's recommendation outcome is by modifying their own user model, triggering changes in the group preferences and resulting in a new set of recommended items. All of the involved preference models (owned one, rest of the member's one and group's one) are accessible by every user, facilitating the negotiation and the discussion, mainly focused on which attributes participants should include in their individual user models, so the group's one is refined.

For a second approach, collaboration has a more relevant place during the process and no individual attributes are defined, but users create and modify the group preference model directly by proposing and voting which attributes should be part of it. Still, they can singly provide an importance level for each attribute that has been selected, whose aggregated values are taken as the significance level of that particular feature, indicating how much it influences the resulting recommendations. Now, the discussion relies on what attributes should be accepted into the model by the group and their significance instead of individually choosing them.

In a last revision, we aimed at simplifying the users interaction by cutting down the steps involved during the attribute negotiation phase. The method is streamlined so the different stages an attribute goes through while creating, proposing and accepting it into the group preference model are now joined in one single step. Users add attributes directly into a shared space where the group preference model is created, without the necessity of accepting them beforehand. There, users only need to specify an importance level for them and the system provides the matching recommendations. Thus, the group discussion is concentrated on a single kind of attribute.

The next sections present this approaches in a more thorough way, together with their respective prototypes and the conclusions we obtained from their studies.

4. Approach 1: Negotiating Individual Preference Profiles

In a first instantiation of the proposed method, we developed a cyclical recommendation process focusing on individual preference elicitation, where the features contained in individual user preference models could be discussed and negotiated by the whole group for influencing the group preference model and, consequently, the recommended items. The details of this method are presented in Ref. [1], which can be summarized as follows:

- Members of a group can create their own individual preference models by selecting the desired item features and ordering them by importance. These individual preferences are publicly accessible by the rest of the group, but only alterable by their owners.
- The system aggregates all the individual preferences to generate a group preference model, used to obtain group recommendations in real time.
- Recommended items are discussed. If consensus about which one to choose cannot be reached, group members can negotiate and modify their individual preference models, initiating the cycle one more time.

The ability to look into other users' preference model, as well as having immediate access to the aggregated model and resulting recommendations, increases the participants' awareness of others' preferences and the effects their own preferences have on the group results. In contrast to a fully automated recommender system, users have a higher level of control over the process and can easily adapt it to their current situational needs and context.

4.1 Prototype 1: Hootle

To test the benefits of our approach, a first prototype, Hootle,

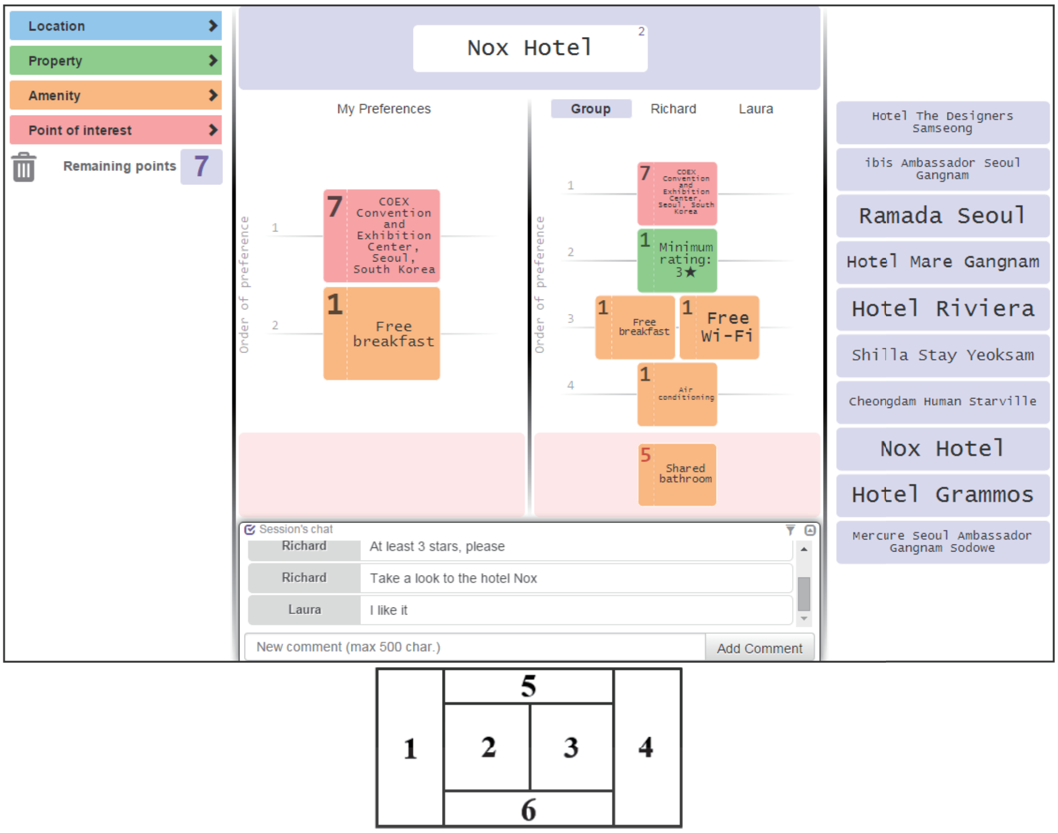


Fig. 1 Different sections of the old interface.

was created. For demonstration purposes, we chose hotel selection for group travel as the application area and used an Expedia dataset consisting of 151.000 hotel entries with descriptive information; the same dataset was used in all three iterations of the development. Despite its focus on the hotel domain, the approach makes use of content-based techniques and is applicable to many different domains, provided the properties of the items to be recommended are available.

Figure 1 depicts the organization of the different areas of the prototype's interface:

(1) `Feature exploration.` A private area for exploring and defining item features by using a set of given filters (e.g., location, facilities or nearby points of interest).

(2) `Individual preferences.` By dragging and dropping the features from the "Feature Exploration" area into this one, users can create a ranked list of features that becomes their individual preference model, where the position of each attribute in the list indicates its importance.

(3) `Group preferences.` The group preference model is dynamically calculated and displayed here every time that a user modifies his or her individual preference model. This area also lets users browse the preferences of the rest of group members.

(4) `Generated Recommendations.` The recommended items are shown in this area, enabling users to access their details and select their preferred ones.

(5) `Proposed items.` The recommended items chosen by users are saved and shared inside this space, so the rest of the group can acknowledge or reject them as a final solution through a voting system.

(6) `Chat.` Here, written discussion is facilitated via chat.

Other minor mechanics were implemented too, such as the addition of a "vetoing sub-area" (bottom of areas 2 and 3, where undesired attributes can be placed), the inclusion of what we called "petitions" (a special kind of comments that specifically ask for the rearrangement of a determined attribute into the group members' individual preference models), an "item approval" system (allowing users to show whether they support a certain recommended item or not) and a "matching score" for every recommended-item/user-model pair (representing how well recommended items suit the individual user preference models).

Regarding the extraction of recommendations, the system takes the group preference model and explores the DB using a content-based filtering method (**Fig. 2**). In content-based filtering, items are described by a set of attributes, which are compared against the preference model of a user (in our case, the aggregation of all individual user models). Because the preference model is created from scratch in each new session, the system is applicable in cold-start situations where no user profile exists yet. Items in the DB are scored depending on how many selected attributes they contain and their rank in the group's model. Once the items have been rated, the system extracts those with the highest scoring. Every time that the group's preference model changes, new recommendations are obtained, enabling real time feedback.

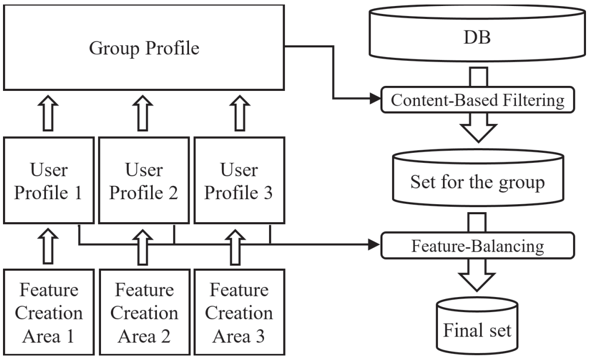


Fig. 2 Recommendation process.

4.2 Feature Balancing

When collecting the items that will be handed to the users as recommendations, it could happen that none of them completely fulfils the group preference model. In the case that only the top rated items were selected, it would be possible that for some of the attributes inside the group preference model not a single matching recommendation is provided, even if they have been highly ranked within some user's individual preference model (due to their average rank still being low). Because the system's raison d'etre is to serve as a tool for discussion and consensus finding within the context of GRS, it makes sense to try to return a well-balanced set of recommendations, allowing these who have chosen less popular attributes to be an active part of the negotiation process. Thus, a further step (which we called feature balancing) is done before sending the matching recommendations to the session's participants, attempting to collect a set of items where there is at least one fitting item per attribute in the preference model.

4.3 Making the Right Decision

Finding a recommendation that matches the group wishes may require several tries. Usually, it will be necessary to move through the different stages of the process in a cyclic and iterative fashion, negotiating the features within the individual preference models to influence the aggregated one and exploring the new matching recommended items once again. When the negotiation and the discussion are the driving force of this changes, with each new iteration the group should get closer to a solution, optimizing the group filters and narrowing down the recommendations.

Nevertheless, even when the process is carried out properly, the criteria for selecting the "right item" may differ from one scenario to another: in some cases, it could be the one that has been accepted by the majority; in others, it could be unacceptable to choose an item that has been rejected by only one member of the group. While a fixed group recommendation strategy might be used, we believe that the system cannot generally resolve such decision problems. Our approach provides tools for preference specification, discussion and acceptance measuring, but it is not possible to talk about the one right solution when dealing with group decision making in a real life situation. Ultimately, it is up to the users to decide whether a recommendation fits their needs or not and to make the final choice.

4.4 Evaluation

A user study was performed to analyse the impact of the cooperative preference elicitation and negotiation tools developed, but also to determine the system's usability and the quality of the resulting recommendations.

4.4.1 Setting and Experimental Tasks

We used the hotel database provided by Expedia with 151,000 entries. For each hotel, a full description and a set of attributes, including property and room amenities (within 360 possibilities), locations (258,426) and points of interest nearby (94,512) were available.

Two different versions of the system were tested. One system version provided the full set of functions described (hereafter version D – Discussion), while the second one was restricted to an individual preference specification and recommended items browsing, with no discussion nor negotiation means enabled (version ND – No Discussion), similar to a conventional group recommender system (therefore, serving as baseline for comparison).

Two types of task scenarios with different levels of complexity were elaborated, a first one for learning the usage of the tools and a second one closer to a real world scenario, where groups had to find a place to stay during the summer vacation.

To prevent participants from complying too quickly with the wishes of other users, we artificially induced different backgrounds and objectives for each group member. For this purpose, we created a set of role cards for the second task that were randomly distributed among the group's members, with the intent of generating conflicts and discussion (e.g., "sport activities", "shopping possibilities", "cultural events", "nature nearby").

4.4.2 Method

48 participants took part in the study (5 males, 43 females, average age of 20.94, 5.018), distributed in groups of different sizes: 4 groups of 3 persons (12), 4 groups of 4 persons (16) and 4 groups of 5 persons (20). One half of the groups of each size worked with the ND version, while the other half ran the D one.

Participants had up to 40 minutes to complete each task (D version was considered completed if consensus was found or the time reached the limit; for the ND version, participants only needed to individually create a preference model they were happy with and, once the whole group had finished, unilaterally choose a recommended item). After completing both tasks, participants were asked to fill in a questionnaire regarding aspects such as the quality of the recommendations or the ease-of-use of the system, using a 1–5 scale. It comprised the SUS items [8] as well as items from two recommender-specific assessment instruments (User experience of RS [17] and ResQue [35]). The recommender-specific items measure the constructs *user-perceived recommendation quality, perceived system effectiveness, interface adequacy, and ease of use*.

4.4.3 Results and Discussion

Members in ND groups were not able to choose the same hotel in a single instance. In two of these cases, some users couldn't even find a hotel that they liked when working on the second task, while all groups with version D were able to choose one unique hotel in both tasks. With respect to the usability, both system versions received a borderline SUS score with no differences be-

Table 1 Some results of the first user evaluation. All the D/ND differences $p > 0.05$, effects of group size were significant.

		No Discussion				Discussion			
		3	4	5	Avg.	3	4	5	Avg.
Overall Satisfaction	m	3.40	3.00	3.70	3.39	4.33	4.00	3.60	3.92
	σ	0.54	1.20	0.48	0.83	0.51	0.53	0.96	0.77
Would recommend it	m	3.20	2.38	3.30	2.96	3.50	3.25	3.30	3.33
	σ	1.30	1.06	0.67	1.02	0.83	0.70	1.06	0.86
Would use it again	m	2.40	2.50	3.10	2.74	3.17	3.13	3.00	3.08
	σ	0.89	0.92	1.10	1.01	0.75	0.99	0.66	0.77
Would use it frequently	m	1.60	1.88	2.30	2.00	2.67	2.75	2.70	2.71
	σ	0.54	0.64	0.67	0.67	0.81	1.04	0.94	0.90
Recommendations were well chosen	m	3.20	3.38	3.80	3.52	4.33	3.38	4.00	3.88
	σ	0.83	0.74	0.78	0.79	0.51	0.74	0.47	0.68

tween them (ND = 68, D = 69).

A 2×3 ANOVA was performed, **Table 1** lists some of the most significant results. The questionnaire results showed a tendency in favour of the D system in the majority of the items (also in the ones not listed here). It seemed reasonable to affirm that group recommender systems certainly can benefit from group discussion and negotiation theory. However, when paying extra attention at the different groups within the D version, the method seemed to be more useful for the smaller ones, who exhibited more satisfaction and willingness to use and recommend the system.

4.5 Lessons Learned

Many of the participants had issues when following the flow of action during the session, mostly due to having too many things happening at the same time. For a big group, discussing single attributes from the group's preference model can quickly become a complicated task, considering that every change a user does in his/her own model will modify the group's model as well, leading to a constant change of on-screen attributes; furthermore, there is an extra effort in browsing each participant's individual model separately that could easily overwhelm an inexperienced user.

Despite the advantages this technique could bring to group recommendations, we concluded that group scalability was a problem in this prototype. It was needed to diminish the complexity of the process, decreasing the sources of information and reworking the preference elicitation mechanism in a way that it is both easy to follow and transparent for all the session's partakers. This issues motivated the modification of the method, matter discussed with more detail in the next section.

5. Approach 2: Negotiating Group Preferences

Based on the findings of the first user study, we developed a revised approach, reported in detail in Ref. [2], mainly aiming at alleviating the problems that arose for larger groups in the first system. Our conclusion from the previous approach was that individually creating preference models and exposing all individual profiles to the group for inspection created a high level of complexity, especially when the number of participants increased and more profiles needed to be observed in order to come to a joint decision. Therefore, modifications especially regarding this aspect seemed necessary. In contrast to the original approach, where the users' individual preference models were explicitly shown, in the revised version users need to collaborate to create the group's model by proposing, filtering and rating attributes in a shared space, keeping the flow of action more simple and transparent even with larger groups. The process is carried out as follows:

- Each participant can individually select the features that they think the recommended items should possess by placing them in a private area.
- Once a feature has been selected, the user may propose it to the rest of the group, and associate a personal relevance score to it.
- By proposing a feature, it becomes visible to the whole group, which will decide whether to accept it as a filter or not by using the provided voting system.
- If the feature is accepted, it becomes an active filter with a given significance, calculated through the aggregation of all the importance levels that each user has assigned to it. A user's personal importance is adjustable at any moment, instantaneously reflecting its impact on the overall significance of a feature and bringing up new recommendations after any change. The set of all the accepted filters and their significance level form the group's preference model.
- Finally, a user is able to highlight specific recommended items and state an opinion (via voting/discussing) about the ones that have been selected by other participants. More features can be proposed, accepted and rated continually, so the recommendations are narrowed down until the group finds an item that satisfies its needs.

As in the previous method, the user's awareness of others' preferences is still increased when compared to normal GRS, due to the possibility to specify the filters' importance individually, having an immediate feedback in the group's model and the recommendations. However, the revised approach now also entailed aspects of critique-based recommenders during the preference elicitation phase, since users could criticize or accept proposed features. In addition, users were able to control the sequence of exposing their preferences at the feature level, which may help in better adapting one's negotiation strategy to the situation at hand.

5.1 Prototype 2: Hootle+

A new, completely redesigned version of the Hootle GRS was implemented, called Hootle+, still making use of a content-based filtering method and the same Expedia hotel database as in the

Fig. 3 Different sections of the interface.

previous one. The remade interface can be seen in **Fig. 3**, comprising the following areas:

(1) `Feature exploration`. A private area for exploring item features by using a set of given filters (e.g., location, facilities or nearby points of interest). It is also possible to provide an importance level together with a short explicative sentence and to specify if the attribute is negative or positive.

(2) `Proposed features`. The attributes that have been proposed are shown into this area, which is shared by all participants. Voting is enabled for each proposed attribute, which can be accepted as a group filter, rejected or vetoed, depending on the results.

(3) `Accepted features`. This area contains the attributes that have been approved (or vetoed) by the group. Together with their specific significance level, these attributes define the group's preference model.

(4) `Recommended items`. The system calculates and displays recommendations into this area. The list is constantly updated in real-time when some group filter is added/removed or its significance changes.

(5) `Selected items`. The recommended items selected by users are placed here, so other participants can see and up-vote or down-vote them.

(6) `Chat`. An area to discuss arbitrary questions that come up during the decision process. It provides the possibility to filter the discussion into threads where specific attributes and items are considered.

In the new prototype, the action has been moved from several individual spaces to one unique shared space where all the participants must collaborate to create the group's preference model at two different but highly intertwined stages: firstly they have to propose and vote the attributes that will be part of the group's model, and secondly, they will rank the accepted attributes to indicate how important they are for the group. Attribute ranking is made by directly assigning individual importance values (in a scale from 1 to 100) that are aggregated instantly. We discarded the old mechanic where attributes where ranked by using ordered lists because it caused many issues regarding information complexity and readability in the first prototype. It is possible to go back and forth between these two stages, proposing new features and removing already accepted ones at any moment. Any other existing functionality (like vetoing attributes or exploring, proposing and voting recommended items) that was already present in the previous prototype, has been implemented in this one too, but adapted for the new method when needed.

Recommendations are generated in a similar way to how it was done before (**Fig. 4**). The system compares the items in the database against the collaboratively created group preference model. Items are rated depending on their significance within the group's model, and the most important ones are extracted. Before displaying them to the users, items are filtered one more time to provide a balanced set of recommendations, this time taking into account the individual importance level that each user has given

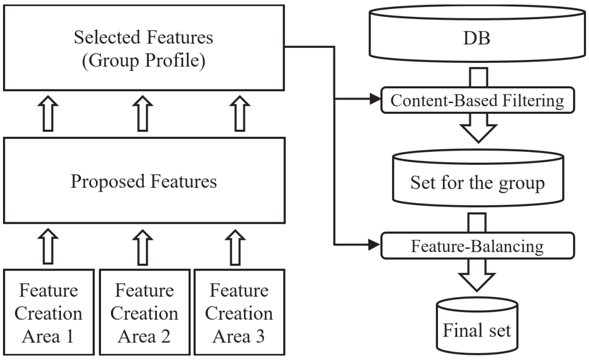


Fig. 4 Recommendation process in the second prototype.

to each feature, instead of using the aggregated values.

5.2 Evaluation

We performed a user study with several groups comprising either three or six users, which is the range of group sizes we expect to occur in real applications. In the user study of the previous system version, we noticed an interesting correlation effect between the group size and its satisfaction, but had groups of three, four and five members, which may have limited the reliability of the results due to the limited range. We thus decided to slightly increase the range and focus on the extreme values. The main objective of this study was to corroborate whether the changes made on the method were of any benefit, for what we needed to determine the usability of the approach and the quality of the resulting recommendations.

5.2.1 Setting and Experimental Tasks

We used the same hotel database provided by Expedia with 151,000 entries and their corresponding descriptions, amenities and information about locations and points of interests.

We prepared two task scenarios with different levels of complexity: in the 'introductory' task, the group was instructed to select a hotel knowing beforehand some common, desired attributes, as well as the location of the hotel; in the 'open' one, only unspecific instructions were given to the group (like finding a hotel to stay during summer vacations).

Like in the precursor study, a set of roles was created and given to participants during the realisation of the second task to promote discussion. A problem detected in the preceding user evaluation was that the roles used were so different from each other that in many cases they created an artificial situation that is not commonly found in real life, where groups that plan to travel together tend to share similar preferences. Thus, for this occasion the roles were simplified and created with shared characteristics:

(1) You love shopping and you are interested in cultural things.
(2) You are interested in cultural things and clubbing.
(3) You love partying every night. During the day, shopping keeps you awake.
(4) You like to spend your time on the beach. When that is not possible, hiking fits well.
(5) You prefer to hike the whole day and do sport related activities.
(6) You are a sport addict and you love the beach.

5.2.2 Method

39 people (22 females, 17 males, average age of 22.63, 3.65) took part in the study, distributed in 5 groups of 3 participants (15) and 4 groups of 6 (24). Since the system is web-based, all users were provided with a normal desktop computer with a display screen of 21" and running the same browser. They sat in a large lab room but were separated from each other and instructed to communicate only via the means provided by the system.

Each group first received a brief introduction to the system and was asked to work on the two decision tasks, always in the order introductory task – open task. Before beginning the second task, they all received randomly one of the role cards. A task was considered complete when the group found consensus (i.e. agreed on a hotel) or the time ran out (25 minutes maximum per task).

After completing both tasks, participants were asked to fill in a questionnaire regarding aspects such as the quality of the recommendations or the ease-of-use of the system (Refs. [8], [17], [35]), same than in the preceding study.

5.2.3 Results

Not all groups were able to find a solution, reaching the time limit for the tasks. For the 3 person groups, agreement was always achieved in contrast to the 6 person groups, where only a 25% of the tasks were completed with consensus regarding the item to select. An average success rate over all sessions of 66% was reached. Despite the low success ratio for the bigger groups, the percentage of agreement among users (participants who selected the same hotel) was 77%, as shown in the objective data listed in **Table 2**. Time needed per task was higher for the 6 people groups, as well as the amount of individual preference changes made per user (importance level, vote selection), but the number of comments written per user in the bigger groups was lower than in 3 people groups. This could mean that participants in bigger groups made a more extensive use of the graphical interface for showing their wishes and opinions to the rest of the group, because relying only in chat communication for transmitting ideas is usually more complicated the more people are writing at the same time. Despite these differences, both group types elaborated preference models with similar sizes.

When compared against the values obtained during the evaluation of the first version of the system, the average size of the preference model created was smaller in the current version than in the older one, where the number of features were easily doubled. Participants made more use of the chat in the old system, possibly explained by the fact that they had less ways to transparently express their opinions (no public voting system for attributes nor significance assignment). Surprisingly, in the earlier version groups where able to find consensus in all the cases, perhaps due to having less time constraints back then than in the new study.

In relation to the usability of the system, it received a SUS score of 65, placing the prototype slightly under the average. An independent-samples t-test was conducted to compare the items of the questionnaire, taking group size as independent variable. While many items did not show a big difference between cases (**Table 3**), some conclusions can be extracted from them. In general, it seems harder for bigger groups to find recommendations

Table 2 Objective results. Lower (LB) and upper (UB) bounds at 95% confidence interval. Last column has the values of the first version of the system, when applicable.

	3 people groups			6 people groups			Avg	Old Version
	m	LB	UB	m	LB	UB	m	m
Time per task (minutes)	13.60	10.18	17.01	17.63	13.8	21.43	15.61	19.9
Preference Model Size	6.10	3.85	8.34	6.38	3.87	8.88	6.23	15
Changes per user	12.33	6.123	18.54	14.56	11.09	18.03	14.35	—
Comments per user	7.16	2.42	11.90	6.41	3.77	9.06	6.92	10.33
Solution found	100%	—	—	25%	—	—	62.5%	100%
Agreement among users	100%	—	—	77%	—	—	88%	—

Table 3 Some results of the evaluation.

	3		6		Avg	
	m	σ	m	σ	m	σ
The recommended items fitted my preferences	4.00	0.50	3.83	1.16	3.88	1.02
I liked the items recommended by the system	3.78	0.83	3.79	0.88	3.79	0.86
It was very easy to find a good solution together	3.78	1.09	2.62	1.31	2.94	1.34
The other team mates agreed my opinion	4.00	0.70	3.29	1.19	3.48	1.12
Even with different opinions we could find a good compromise	4.44	0.73	3.46	1.06	3.73	1.06
I can make a better choice with the system	3.78	0.97	3.96	1.2	3.91	1.18
I can find a solution in less time using the system	3.56	1.33	4.04	1.08	3.91	1.15
I think the program is easy to use	3.67	0.87	3.46	1.06	3.52	1.00
I think the functions in this program are well integrated	3.56	0.88	4.00	0.72	3.88	0.78
In general, I am satisfied with the system	3.56	1.13	4.33	0.96	3.76	1.00

*Significant ($p < 0.05$)

that match the participants' individual wishes and to agree with the rest of the members, which is a logical consequence of the group size's increase. Interesting is the fact that the groups of 6 are in general more satisfied with the tool than the smaller groups, despite being easier for the latter to find a solution through consensus.

Regarding the old system, the average satisfaction was of 3.92 ($\sigma - 0.77$), surpassing the one obtained in the new version; however, taking a closer look to the results collected for each group size (3 persons: m $-$ 4.33, $\sigma - 0.51$; 4 persons: m $-$ 4.00, $\sigma - 0.53$; 5 persons: m $-$ 3.60, $\sigma - 0.96$) it is apparent that the satisfaction tended to be inversely proportional to the number or participants, an issue not encountered in the more recent user study. This finding supports our hypothesis that the revised system scales better with group size, i.e. it also supports larger groups well.

5.2.4 Discussion

The outcome of the evaluation indicates that some of the issues found during the first user study have been lessened, specifically the one related with how well the system scales up with the group size. Even if having bigger groups increases the complexity of the decision-making process, the results point to a greater satisfaction and sense of helpfulness when using the system. This is more noticeable when one looks to the preference model size, which is almost the same through group sizes indicating that users limited the number of preferences expressed in a well-considered manner in order to facilitate consensus finding. The low ratio of solutions found for the 6 people groups could be explained as a consequence of limiting the time to finishing a task to only 25 minutes, but further research may be needed in order to obtain some final conclusions. In a real world situation, where the time span for finding a solution in a non-collocated group setting could be days or even weeks, and where individual preferences may tend to be more homogeneous without artificially inserting roles, a higher success ratio would be expected.

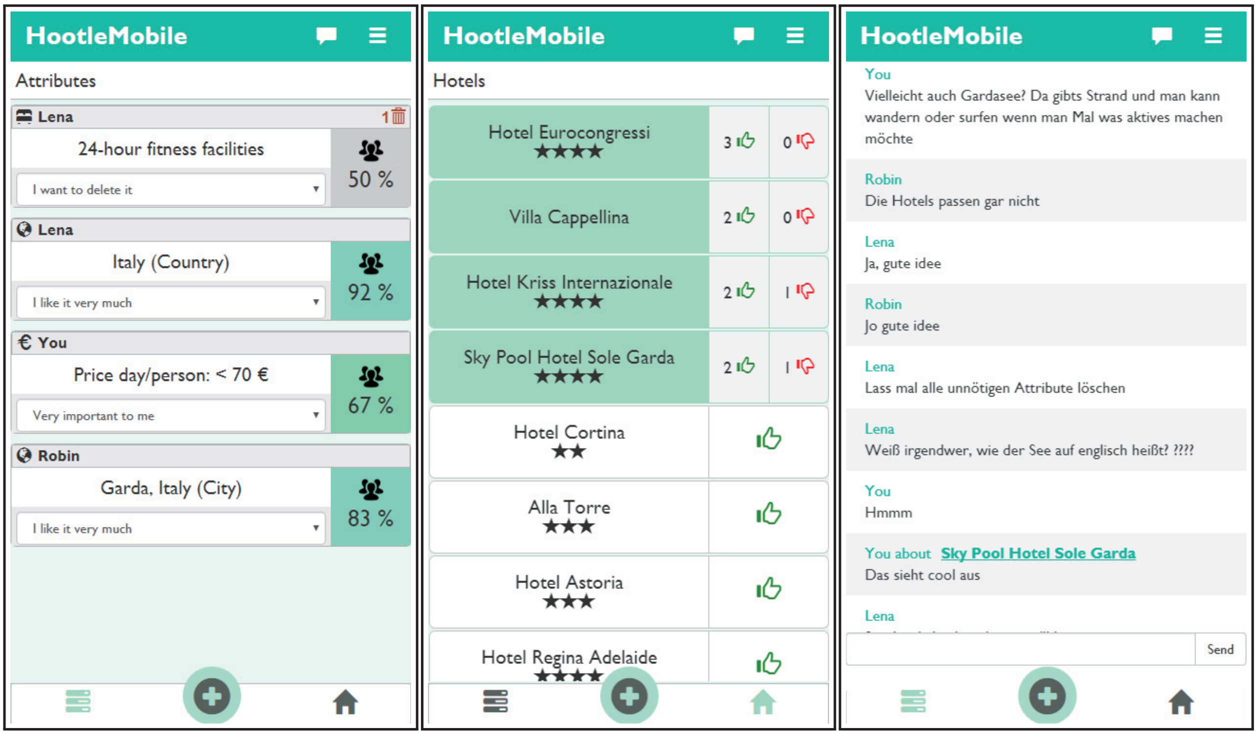
6. Approach 3: Directly Exposing Preferences to the Group

Even though the results of the last evaluation suggest that the revised method improves the scalability for larger groups, there were still issues regarding the complexity of the user interface which resulted in a relatively low SUS score. A major concern was related to the strict separation between the private space for expressing one's own preferences and the group space accessible by all participants. In this section, we report on a third version of the system in which the separation between private and public area was removed. While we expected this further simplification of the process to improve the overall usability, the new design was at the same time more suitable for a mobile version of the system.

6.1 Modified Preference Proposition

In the previously discussed version of the method, users had to create their individually selected attributes inside a private area. Once they were sure about a desired feature they could propose it to the group, but it would only become part of the group's preference model following an approval, where it could be finally rated. While this approach has the advantage of allowing users to reflect on their own preferences before exposing them to the group and the extra filter the attributes had to go through before being part of the group's model is useful in terms of keeping the model low on attribute number, it also introduced an additional step in the process that was considered complex by some of the participants. Furthermore, in a mobile device, where screen space is scarce, having several steps to create a group model would require distributing them over several screens which leads to additional navigation effort, making the interface more cumbersome for users. For these reasons, we modified the process in the following way:

(1) Each participant can directly add the features that recommended items should contain in a shared space, where they



(1) Features (2) Recommendations (3) Chat

Fig. 5 Areas of Hootle Mobile.

are visible for the whole group.

(2) Members of the group can assign an importance level to any of the features in the shared space without the need of approving them first. The mean of the given importance levels is used as the feature's significance.

(3) Features with the highest significance levels become part of the group's preference model and are used to calculate the recommendations. Every time that a user changes the individual importance level given to an attribute, new recommendations are calculated too.

(4) Users are able to highlight specific recommended items and discuss about them. More features can be consequentially added and rated, so the recommendations are narrowed down until a suitable item is found.

With these changes, not only the process has been shortened, but also its representation has been simplified because only two main areas are needed: one for adding/rating features and one for displaying recommendations.

6.2 Prototype 3: Hootle Mobile

The new prototype, Hootle Mobile, employs the same hotel database with 151.000 hotel entries that was used by the other two previous versions. It is still web-based like its predecessors, but for the sake of making it compatible with mobile devices the working space has been split into three different areas, each one of them filling the whole visualization area (accommodating it for small screens), as opposed to the older prototypes where all the relevant information is displayed at once. **Figure 5** shows them:

(1) `Features`. The area where the group's preference model is defined. Users add the attributes they like here (without the need of defining them first in a private area), so they can be rated by the rest of the group members and used by the system for calculating recommendations.

(2) `Recommendations`. Items recommended by the system are displayed in this area, from where the users can highlight those that they like and propose them to the group.

(3) `Chat`. A space where group members can share their thought about the picked attributes or the recommended items.

Most of the functionality offered by previous prototypes has been included in Hootle Mobile too, excepting those application's features that were found seldom used during the two previous studies or the ones that would not perform very well in mobile devices. For instance, negative attributes were removed due to being almost completely ignored by users and the number of vetoed attributes that a single user can specify has been limited to one. Additionally, the new system includes a tutorial, which was not present in the preceding ones.

Regarding the recommendation generation process, the 1–100 importance level scale has been translated to five not numerical options. The numerical scale provides more freedom when deciding the importance level, but having several attributes with only little difference in their importance levels was not very useful for calculating the recommendations, while the five options method makes the decision of which one to assign more relevant (less options, but their values are more distant). Besides, instead of accepting proposed features to make them part of the group's preference model (or to become a filter, as they were called in the previous iteration) now only the ones with the highest signif-

Table 4 Objective results. Lower (LB) and upper (UB) bounds at 95% confidence interval.

	3 people groups			6 people groups			Avg
	m	LB	UB	m	LB	UB	m
Time per task (minutes)	12.66	9.66	16.66	12.16	7.84	16.48	12.41
Preference Model Size	4.89	2.99	6.79	5.83	4.03	7.64	5.36
Comments per user	7.33	5.19	9.41	4.07	1.60	6.53	5.70

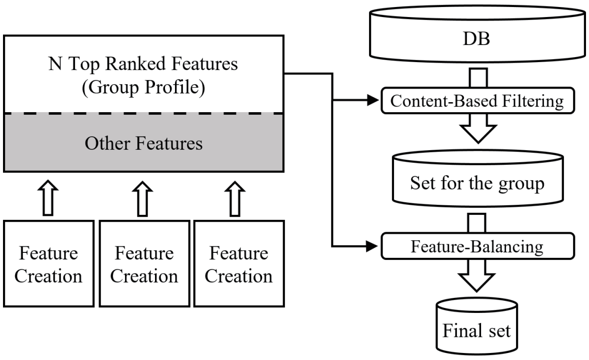


Fig. 6 Recommendation process in the third prototype.

icance levels will be used for calculating the recommendations, removing one extra step in the recommendation process (**Fig. 6**).

Users are still aware of other's preferences, being able to access information regarding how the rest of participants have rated each attribute (importance level) and recommended item. Thus, the negotiation happens at two levels: directly over the individually selected attributes or the chosen recommended items (in which each user can express a personal rating and see the ratings the rest of users have assigned to them) and through chat (where they can discuss the group's preference model and its outcome).

6.3 Evaluation

A new user study was conducted to verify the legitimacy of the most recent changes, supporting our hypothesis that the private space (and by extension, the extra steps that it involved during the creation of the group's preference model) is not necessary and removing it from the process will have a positive impact, alleviating the overall complexity of the system.

We tried to have a similar set-up to the previous study, for what we gathered groups of three and six participants. Since the benefits of using our approach when compared against traditional methods were already explored in the previous evaluations, there was no need of using a limited prototype that worked as a base line for this study and all the groups worked with the exact same version of Hootle Mobile.

6.3.1 Setting

The Expedia hotel database was used again, containing 151,000 hotel entries and their correspondent descriptions, property and room amenities, locations and points of interests. Groups had to work through two different task scenarios: an introductory one, where participants were asked to find a Hotel to stay during a conference in Berlin, breakfast included; and an open one, where the groups had to discuss where to go for the summer vacations, with no given restrictions of any kind.

As in previous occasions, participants were assigned different roles with the objective of creating conflicts, avoiding situations where they could comply to easily one with each other. This study used the exact same roles that the ones in the Hootle+ evaluation.

6.3.2 Method

The study included a total of 42 persons (22 females, 20 males, average age of 27.33, 9.16), divided in 4 groups of 6 participants and 6 groups of 3. Since the prototype has been designed to run under mobile devices, each participant received one with the system already running on it. They sat in the same room, one group at a time, with instructions of not make use of any other means of communication than the ones provided by the recommender system. Then, participants were told to go through the tutorial, with no further explanation about how to use the tool. When all members of a group were finished, roles were randomly designated among them, who could now start with both the introductory and the open task (in this order), for what they had a time limit of 25 minutes per task.

When both tasks were completed (or the time limit reached), participants had to fill in the same questionnaire that was used for the previous study, thus allowing us to compare the results. The questionnaire included SUS items [8] together with items from two recommender-specific assessment instruments (User experience of RS [17] and ResQue [35]) that measure the constructs *user-perceived recommendation quality, perceived system effectiveness, interface adequacy, and ease of use*.

6.3.3 Results

All the groups were able to find a solution all their members agreed with within the given time. **Table 4** contains some objective results collected during the study, showing slightly lower numbers in terms of time per task, group's model size and comments per user (the number of changes a user did was not recorded during this evaluation).

A two way ANOVA test for comparing Hootle+ and Hootle Mobile has been performed, whose significant results are listed in **Table 5**. The questionnaire's results were always better in the newest prototype, where most of the significant values are found in items regarding complexity, aesthetics and willingness to use the system again. In any case, no item performed worse in Hootle Mobile than in Hootle+. The SUS score was significantly better too, with a final value of 82 against the 65 obtained by Hootle+.

6.3.4 Discussion

Results of the evaluation denote that the changes made in the method were actually an improvement, confirming our initial expectations. Removing the private area does not seem to have any drawbacks in the recommendation process, and the streamlined method together with the consequently simplified user interface have had a positive impact on the user experience. Regarding the objective results, reducing the number of steps might be the cause of the observed time per task decrement, while the lower number of comments per user could be explained by the usage of

Table 5 Two-way ANOVA test significant results at $p < 0.001$.

		Hootle+			Hootle Mobile		
		3	6	Avg	3	6	Avg
The layout of this recommender system interface is attractive	m	2.44	3.21	3.00	4.06	3.96	4.00
	σ	1.236	1.062	1.15	0.83	1.11	0.99
I became familiar with this recommender system very quickly	m	3.22	3.38	3.33	4.65	4.43	4.53
	σ	1.09	1.17	1.13	0.60	0.79	0.72
Overall, I am satisfied with this recommender system	m	3.56	3.83	3.76	4.41	4.39	4.40
	σ	1.13	0.96	1.00	0.62	0.72	0.67
I will use this recommender again	m	3.56	3.50	3.52	4.47	4.43	4.45
	σ	1.24	1.29	1.25	0.72	0.84	0.78
I will use this recommender frequently	m	3.11	2.63	2.76	4.06	3.96	4.00
	σ	1.69	1.14	1.30	1.03	0.92	0.96
I will tell my friends about this recommender	m	3.44	3.75	3.67	4.47	4.48	4.48
	σ	1.51	1.11	1.22	0.72	0.73	0.72
I found the system very cumbersome to use	m	2.44	2.83	2.73	1.65	1.70	1.68
	σ	1.13	1.24	1.21	0.79	0.70	0.73

mobile devices, considering that writing on a touch-screen might be harder than doing it on a physical keyboard.

7. Comparison of the Approaches Based on an Initial GRS Model

Based on the experience gained with the three approaches developed and the empirical results we can more clearly distinguish the different aspects and phases of a group decision process supported by a GRS. As an initial model of such processes we suggest to distinguish the following phases each of which also has cognitive correlates and can be supported by specific system functionality:

(1) Users make themselves aware of their preferences, express them, reflect on them and potentially adapt them either based on their own insight or through interaction with other group members.

(2) Users reveal and communicate their preferences to other group members or the whole group, either as complete preference profiles or as single feature preferences.

(3) Group members discuss, criticize, or weight the individual preferences or the group model as a whole, possibly involving voting mechanisms to decide on the acceptability of individual preferences.

(4) Group members weight, criticize or vote the resulting recommendations, converging on a joint decision.

The three approaches described in this paper each focus on these phases to a different extent. Each of them strikes a different balance between private preference spaces and public spaces where other users can see, criticize and discuss the individual or group preferences. As a consequence, the number of interaction steps an individual user needs to take in the overall process differs. Hootle Mobile directly exposes each feature selected by a user to the whole group which results in an increased efficiency in comparison to the other approaches. Not surprisingly, the usability related metrics (e.g., as measured by SUS) are significantly more positive than in the first two approaches. Also the overall satisfaction with the system and the recommendations given were more positive. Scalability is also an important criterion when decision making in larger groups is to be supported. Here, the complexity of the system increases with the amount of information about individual preferences presented. Especially for the first approach, which showed all individual preference profiles to the whole group, problems were found in this aspect. Again, the direct presentation of each preference in the group space, as applied in Hootle Mobile has shown to be advantageous. In terms of recommendation quality, Hootle Mobile also received higher ratings than the other versions, especially for larger groups. This difference was not present for small groups between Hootle+ and Hootle Mobile.

Overall, the simplified process implemented in Hootle Mobile resulted in better scores for usability-related criteria, scalability for larger group sizes, and also perceived recommendation quality in the case of larger groups. It has to be noted, however, that these results were obtained only for a single recommendation domain (hotels) which may have influenced the negotiation strategy used by the group members. In general, this domain, especially in experimental conditions, tends to lead to group decisions that are not very controversial. There are other domains, however, that involve more risk for the individual which may lead to different negotiation strategies. In the field of negotiation research for example [10], is has been shown that the sequence in which a participant reveals his or her preferences or offers to the other stakeholders may influence the success of the negotiation. For such high-risk negotiations, for example the purchase of high-price products or investment decision, it may be more appropriate for individual to first externalize their preferences in a private space before deciding which preference to communicate to the group. In such contexts, the need for system support may be distributed differently over the four phases described above, focusing more strongly on phase 1, as was the case in Hootle and Hootle+. In general, however, our studies provide evidence that the less complex method of directly submitting individual preferences to the group for discussion and voting is more usable and acceptable. Nonetheless, effectively supporting the different phases of the model outlined above is an area for further investigation.

8. Conclusion and Outlook

In this paper, we have presented an approach to group recommender systems, investigating it by means of two systems versions that we empirically evaluated. The method enables collaborative preference elicitation on the fly, avoiding a cold-start situation and providing more control during the recommendation process. The system supports the negotiation and the discussion during the preference elicitation and item selection phases. Participants can freely define and propose features, adding them to a shared pool of attributes where the group will collaboratively select those to be part of the group preference model. Once the attributes are extracted, users are able to individually assign an importance level to each one of them and the system calculates their significance to the group. Recommendations are then generated after the given group preference model and will be recalculated each time that it changes. Recommendations are shown to the group members, letting them select and discuss about those that they like, or redefine the group preference model to obtain new recommended items.

The technique herein described provides higher flexibility and awareness than the fixed strategies typically used in group recommenders. Since preferences and matching recommendations are always visible, participants' awareness of individual and group views and of the effects of their preference settings is increased.

Based on prior work and the ideas described above, a new prototype version of our hotel group recommender, Hootle Mobile, was developed. The results of the user study we conducted show that the new prototype performs significantly better that the ones created in previous iterations, providing a simplified method when maintaining all the capabilities of its predecessors.

Testing the method with real groups is still a pending subject, since their feedback would be a solution to the problem inherent to the use of artificial roles during the test sessions. Furthermore, with enough user data, it would be possible to create predictions based on what other groups had chosen in the past by using a collaborative filtering approach, providing an initial set of desired attributes and further lightening the feature selection stage; another possibility in that regard would be to exclude recommended items (or highlight them) similar to those that were rejected (or accepted) in past sessions. Finally, it is also in our scope to further develop a model for negotiation-based group recommending, which is outlined in a initial form in this paper.

Acknowledgments We are grateful to Vanessa Vogel for supporting the development and evaluation of Hootle Mobile.

References

[1] Álvarez Márquez, J.O. and Ziegler, J.: Preference elicitation and negotiation in a group recommender system, *Human-Computer Interaction*, pp.20–37, Springer (2015).

[2] Álvarez Márquez, J.O. and Ziegler, J.: Hootle+: A Group Recommender System Supporting Preference Negotiation, *CYTED-RITOS International Workshop on Groupware*, pp.151–166, Springer (2016).

[3] Ardissono, L., Goy, A., Petrone, G. and Segnan, M.: A multi-agent infrastructure for developing personalized web-based systems, *ACM Trans. Internet Technology* (*TOIT*), Vol.5, No.1, pp.47–69 (2005).

[4] Ardissono, L., Goy, A., Petrone, G., Segnan, M. and Torasso, P.: Intrigue: Personalized recommendation of tourist attractions for desktop and hand held devices, *Applied Artificial Intelligence*, Vol.17, No.8–9, pp.687–714 (2003).

[5] Beckmann, C. and Gross, T.: Towards a group recommender process model for ad-hoc groups and on-demand recommendations, *Proc. 16th ACM International Conference on Supporting Group Work*, pp.329–330, ACM (2010).

[6] Bekkerman, P., Kraus, S. and Ricci, F.: Applying cooperative negotiation methodology to group recommendation problem, *Proc. Workshop on Recommender Systems in 17th European Conference on Artificial Intelligence* (*ECAI 2006*), pp.72–75 (2006).

[7] Boratto, L. and Carta, S.: State-of-the-art in group recommendation and new approaches for automatic identification of groups, *Information Retrieval and Mining in Distributed Environments*, pp.1–20, Springer (2010).

[8] Brooke, J. et al.: SUS-A quick and dirty usability scale, *Usability Evaluation in Industry*, Vol.189, No.194, pp.4–7 (1996).

[9] Doodle, available from ⟨http://www.doodle.com⟩.

[10] Fatima, S. and Wooldridge, M.: Optimal agendas for sequential negotiations, *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pp.41–55, Springer (2014).

[11] Hartnett, T.: *Consensus-oriented decision-making: The codm model for facilitating groups to widespread agreement*, New Society Publishers (2011).

[12] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T.: Evaluating collaborative filtering recommender systems, *ACM Trans. Inf. Syst.* (*TOIS*), Vol.22, No.1, pp.5–53 (2004).

[13] Hill, W., Stead, L., Rosenstein, M. and Furnas, G.: Recommending and evaluating choices in a virtual community of use, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.194–201, ACM Press/Addison-Wesley Publishing Co. (1995).

[14] Jameson, A.: More than the sum of its members: challenges for group recommender systems, *Proc. Working Conference on Advanced Visual Interfaces*, pp.48–54, ACM (2004).

[15] Jameson, A., Baldes, S. and Kleinbauer, T.: Two methods for enhancing mutual awareness in a group recommender system, *Proc. Working Conference on Advanced Visual Interfaces*, pp.447–449, ACM (2004).

[16] Jameson, A. and Smyth, B.: Recommendation to groups, *The adaptive web*, pp.596–627 (2007).

[17] Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H. and Newell, C.: Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction*, Vol.22, No.4-5, pp.441–504 (2012).

[18] Kompan, M. and Bielikova, M.: Group Recommendations: Survey and Perspectives, *Computing and Informatics*, Vol.33, No.2, pp.446–476 (2014).

[19] Lieberman, H., Van Dyke, N. and Vivacqua, A.: Let's browse: A collaborative browsing agent, *Knowledge-Based Systems*, Vol.12, No.8, pp.427–431 (1999).

[20] Liu, X., Tian, Y., Ye, M. and Lee, W.-C.: Exploring personal impact for group recommendation, *Proc. 21st ACM International Conference on Information and Knowledge Management*, pp.674–683, ACM (2012).

[21] Loepp, B., Hussein, T. and Ziegler, J.: Choice-based preference elicitation for collaborative filtering recommender systems, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.3085–3094, ACM (2014).

[22] Masthoff, J.: Group modeling: Selecting a sequence of television items to suit a group of viewers, *Personalized Digital Television*, pp.93–141, Springer (2004).

[23] McCarthy, J.F. and Anagnost, T.D.: MusicFX: An arbiter of group preferences for computer supported collaborative workouts, *Proc. 1998 ACM Conference on Computer Supported Cooperative Work*, pp.363–372, ACM (1998).

[24] McCarthy, K., McGinty, L. and Smyth, B.: Case-based group recommendation: Compromising for success, *International Conference on Case-Based Reasoning*, pp.299–313, Springer (2007).

[25] McCarthy, K., McGinty, L., Smyth, B. and Salamó, M.: The needs of the many: A case-based group recommender system, *European Conference on Case-Based Reasoning*, pp.196–210, Springer (2006).

[26] McCarthy, K., McGinty, L., Smyth, B. and Salamó, M.: Social interaction in the cats group recommender, *Workshop on the Social Navigation and Community Based Adaptation Technologies* (2006).

[27] McCarthy, K., Salamó, M., Coyle, L., McGinty, L., Smyth, B. and Nixon, P.: CATS: A Synchronous Approach to Collaborative Group Recommendation, *FLAIRS Conference*, Vol.2006, pp.86–91 (2006).

[28] McGrath, J.E. and Berdahl, J.L.: Groups, technology, and time, *Theory and Research on Small Groups*, pp.205–228, Springer (2002).

[29] Nguyen, T.N. and Ricci, F.: Dynamic elicitation of user preferences in a chat-based group recommender system, *Proc. Symposium on Applied Computing*, pp.1685–1692, ACM (2017).

[30] Nunamaker Jr, J.F., Briggs, R.O., Mittleman, D.D., Vogel, D.R. and Pierre, B.A.: Lessons from a dozen years of group support systems research: A discussion of lab and field findings, *Journal of Management Information Systems*, Vol.13, No.3, pp.163–207 (1996).

[31] O'connor, M., Cosley, D., Konstan, J.A. and Riedl, J.: PolyLens: A recommender system for groups of users, *ECSCW 2001*, pp.199–218, Springer (2001).

[32] Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R. and Yin, F.: Empirical analysis of the impact of recommender systems on sales, *Journal of Management Information Systems*, Vol.27, No.2, pp.159–

188 (2010).
[33] Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.-P. and Jonker, C.M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process, *User Modeling and User-Adapted Interaction*, Vol.22, No.4-5, pp.357–397 (2012).
[34] Pu, P. and Chen, L.: User-involved preference elicitation for product search and recommender systems, *AI magazine*, Vol.29, No.4, p.93 (2009).
[35] Pu, P., Chen, L. and Hu, R.: A user-centric evaluation framework for recommender systems, *Proc. 5th ACM Conference on Recommender Systems*, pp.157–164, ACM (2011).
[36] Ricci, F., Rokach, L., Shapira, B. and Kantor, P.: *Recommender Systems Handbook*, chapter Introduction to Recommender Systems Handbook, pp.1–35, Springer (2010).
[37] Saaty, T.L.: *Fundamentals of decision making and priority theory with the analytic hierarchy process*, Vol.6, Rws Publications (2000).
[38] Stettinger, M.: Choicla: Towards domain-independent decision support for groups of users, *Proc. 8th ACM Conference on Recommender Systems*, pp.425–428, ACM (2014).
[39] Walther, J.B.: Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction, *Communication Research*, Vol.23, No.1, pp.3–43 (1996).

Jesús Omar Álvarez Márquez received his M.S. degree in Computer Science at the University of Granada in 2014. During the same year, he joined the Interactive Systems Research Group at the University of Duisburg-Essen, where he has been working since then as a researcher in the field of recommender systems and, more recently, augmented reality.

Jürgen Ziegler is a full professor in the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen where he directs the Interactive Systems Research Group. Prior to joining the University, he was head of the Competence Center for Software Technology and Interactive Systems at the Fraunhofer IAO in Stuttgart. He holds a diploma degree in electrical engineering and biocybernetics from the University of Karlsruhe and a doctoral degree from the University of Stuttgart. His main research interests lie in the areas of human-computer interaction, intelligent user interfaces, recommender systems and information visualization. Current projects of his group focus on interactive techniques for recommender systems, interfaces and visualizations for semantic data, social media, and personal health applications.

Paper IV. Augmented Reality Based Recommending in the Physical World

Reproduced with permission from the authors: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Augmented Reality Based Recommending in the Physical World*. In *Mensch und Computer 2018-Workshopband*, Gesellschaft für Informatik eV.

https://doi.org/10.18420/muc2018-ws07-0461

Augmented Reality Based Recommending in the Physical World

Jesús Omar Álvarez Márquez[1], Jürgen Ziegler[1]

Interactive Systems, University of Duisburg-Essen[1]

jesus.alvarez-marquez@uni-due.de, juergen.ziegler@uni-due.de

Abstract

Recommender systems have received the attention of the scientific community for a long time now and they have become a daily tool for internet users. Nonetheless, they are not commonly applied to physical settings, where having access to recommendations could be of great benefit, specially when combined with item comparison capabilities. Due to the latest augmented reality technology advances, it is possible to bring these concepts together. An intuitive action like visually comparing two products could be enhanced by 3D cues and suggestions. In such terms, we discuss the possibilities to improve the item exploration and decision-making stages of the recommending process by providing item comparison supported by 3D augmentations, offering a novel contribution to both augmented reality and recommender systems domains.

1 Introduction

Recommender systems (RS) have become an everyday tool that most internet users know and benefit from (Ricci et al., 2011). They cover a wide range of domains due to their proven usefulness and the extensive research behind them. Furthermore, popular websites and applications offer them as a main part of their services, many of which could not be conceived without them (e.g. Amazon, Trivago). Effectively communicating to the user the reasons behind a given recommendation has proven to be crucial for increasing the system's transparency and trustworthiness (Sinha and Swearingen, 2002). Many methods to enhance transparency have been researched, relying mostly on textual explanations or the way items are presented (Tintarev and Masthoff, 2007). In conjunction to RS, it is also common for online retailers to offer product comparison tools to help users during the decision-making stage.

RS have been largely used in digital settings, but they are rarely applied to real world contexts, despite their potential to be equally valuable when dealing with physical objects. Nonetheless,

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2018 in
R. Dachselt, G. Weber (Hrsg.):
Mensch und Computer 2018 – Workshopband, 02.–05. September 2018, Dresden.
 https://doi.org/10.18420/muc2018-ws07-0461

recent advances in augmented reality (AR) technology allow new interaction methods, bringing opportunities to employ RS theory to physical world situations, where recommendations and the real world objects they concern are presented together in a shared space. How these recommendations are shown and what interaction methods should be used to manipulate them are still open research questions.

When in a physical store context, it is particularly interesting to observe how clients behave if no external information source is at hand (e.g. RS, expert opinions or user ratings), so that the customer must rely on what it is provided by the products themselves (e.g. their appearance or a technical data sheet next to them). In such situation, customers tend to focus on product characteristics, in a decision-making process that requires comparing attributes of different products against each other or against the client's own preferences (Lancaster, 1966). Comparison is one of the most basic cognitive activities and plays an important role in understanding, discovering and evaluating our surroundings (Gentner and Medina, 1997). Nonetheless, retaining product characteristics can be a big constraint when comparing several items (specially when they are not side by side), issue that is accentuated by the limitations of short-term visual memory (Alvarez and Cavanagh, 2004). It has been proven that AR alleviates the mental workload of retaining information by eliminating short-term memory demands by using spatial superimposition (Tang et al., 2003). When combined with a RS, this approach would support the recommending process during the item exploration and decision making stages, improving the comparison action and enhancing the way a user inspects products and their disparities/similarities.

In this paper we present our ongoing research regarding AR supported comparisons in the field of RS, where previous work is introduced and further research discussed.

2 Related Work

AR has received a great amount of attention lately, mostly due to its recent consolidation as an approachable technological choice (Chatzopoulos et al., 2017). In a few cases, RS and AR have been coupled already for product recommendations in brick and mortar stores. Examples of it are PromoPad (Zhu and Owen, 2008), which deepens in the concept of dynamic product contextualization to provide suggestions, and PHARA (Gutiérrez et al., 2017), where an AR system oriented to promote the adoption of healthy food buying behaviours is presented.

Supporting product comparison is a common feature in online retailers, where the characteristics of different products are shown side by side. In the field of RS, critiquing-based recommenders allow users to receive new recommendations by modifying specific feature values of the current, given ones, thus performing a direct comparison (e.g. a film with more action, a car with less gas consumption). Interesting in terms of visualization, Zhang et al. (2008) studies the benefits of using a visual interface which presents critiques of several items at once by displaying icons instead of text.

To the best of our knowledge, visually expressing feature differences and/or similarities of two or more physical objects has not been studied in conjunction to RS and AR yet, although visual comparison research made in other areas might serve as starting point, like studies addressing the comparison of graphs (Gleicher et al., 2011) or maps (G. L. Andrienko and N. V. Andrienko, 1999). In Tominski et al. (2012) a system that supports the comparison of information printed on paper is described, reporting the benefits of using natural interaction methods.

3 Combining AR and RS: Research Status

We aim to study the benefits of providing recommendations supported by AR in a physical store situation, being of great importance to investigate how to convey and interact with product information (whatever it may be) in the virtual world in a comprehensible, natural manner.

In the following, we first present an early approach to our research, where we explore the feasibility to use a virtual advisor that guides the user and provides insight on why products are recommended. After reviewing the lessons learned, a new research direction targeting product comparison visualization is discussed.

3.1 The Initial Study: Product Explanation + Virtual Advisor

In a first attempt to use augmented reality in the field of recommender systems, we developed an application running under Microsoft's HoloLens that is able to recognize a number of physical printers and provide recommendations after collecting the customer's preferences. Multimodal, natural interaction was a priority, enabling natural language recognition (via Google's Dialogflow), selection through gaze and air tapping, as well offering feedback in the way of 3D augmentations, text and text-to-speech. A main focus of the research was to explore product explanation through AR in RS, accomplished by the usage of an embodied virtual advisor (Fig. 1) and virtual augmentations of the products (Fig. 2).

Figure 1: Virtual advisor.

The virtual advisor provides guidance through the buying process, giving under request information relative to the products (e.g. price, availability of features). It also gives instructions

Figure 2: Using augmentations to explain how to refill a printer with paper.

about how to operate a printer (e.g. how to change a cartridge) while at the same time augmentations of the products are shown, visually enforcing what the advisor says. Product explanation is meant to support the buying decision process, being useful in situations where the way a specific operation is conducted is relevant for the client's choice (e.g. finding a product that is controlled in a similar fashion to a previously owned one).

If asked to, the advisor will give content-based recommendations, for which the preference elicitation is carried out in a conversational manner through a number of questions (e.g. are you looking for a printer for your home or your office?). Once a recommendation is given, a user can inquire about the reasons behind it, for what the advisor will expose how the chosen printer matches the specified requirements. If the user is not convinced, it is still possible to critique some of the printer's features to receive a new recommendation.

3.1.1 Outcome

We conducted a small user study (N = 15) from where we soon realised the existence of usability issues. Two important lesson that we learnt were that:

- **The number of information sources should be kept as low as possible, fitting on the screen and anchored to real world.** Splitting the action and information sources proved to be harmful to the experience. The virtual avatar and the information overlay displayed on the printers were not visible at the same time, breaking the immersion and disorienting the user, who in many cases did not know where to focus the attention. Furthermore, the avatar position was lost quite often, due to the lack of a physical anchor.
- **The use of AR technology only makes sense when it adds something unique that cannot be reproduced by any other alternative means.** Keeping consumers' fidelity and willingness to use AR solutions has been referred as an issue already (Chatzopoulos et al., 2017; Hopp and Gangadharbatla, 2016), the main causes being that a) the sense of novelty fades away quickly and b) the existence of other methods that provide a similar service without causing the physical fatigue of holding a camera or wearing special

equipment. Specially in the context of e-commerce, there are examples of mobile applications capable of recognizing products and giving recommendations in a similar setting than the one presented here, which means that in most cases users will try an AR approach out of curiosity, but will not stick to it.

3.2 Current Work

Because of the findings of the first study, the focus of our current research has taken a slightly different direction. While overlaying usage explanation on products is something that cannot be achieved by any other technology aside from AR (bringing an added value per se), our goal is far beyond providing such explanations and the effort of creating them should be minimized. Nonetheless, being able to see differences between products (respecting their usage or any other matter) happens to be very useful knowledge when deciding which one to buy. Following this idea, the research now pursues the goal to investigate how to use AR to directly perceive discrepancies among physical products that are not obvious at plain sight or that require to be consulted on a separate information source to know them. The advantages of using an embodied virtual advisor in this scenario remains an open question, although natural language recognition seems to be a valuable feature. Ultimately, these ideas will still be built on top of a RS, improving item exploration and helping during the decision making stage. In addition, the new concept brings to the light several new questions:

- **What kind of data is useful for a comparison?**
  Traditionally, websites that provide comparing tools simply list all their attributes side by side. AR is a more powerful communication medium, but it has its own restrictions. It is critical to use the right visualization means to report a comparison (text, highlighting parts of the object, animations, navigation aids) while at the same time avoiding to overcrowd the view with too many information sources. Filtering down what to show (also taking into account the user's preferences) and when to do it gains greater significance.

- **How to visualize a comparison of an out-of-sight object?**
  A system that aspires to show dissimilarities among two or more physical objects will have to deal with the fact they will not always be on-screen. Therefore, it is fundamental to study how to keep track of what is being compared and how to translate the comparison to the user in a comprehensible manner.

- **How should interaction be carried out?**
  Studying how humans intuitively behave when comparing objects could be beneficial in the creation of interaction methods that feel effortless and natural. As an example, when using a platform that allows free hand movement (like a head mounted display) the comparison tool could recognize when a user is holding a product to take a closer look to it, consequently augmenting it to display significant data (Fig. 3).

- **How to identify characteristics of suitable product domains?**
  Finding a domain where these concepts perform well might not be an easy task. Small, full of details an easily distinguishable objects are preferred. Also, expert knowledge and

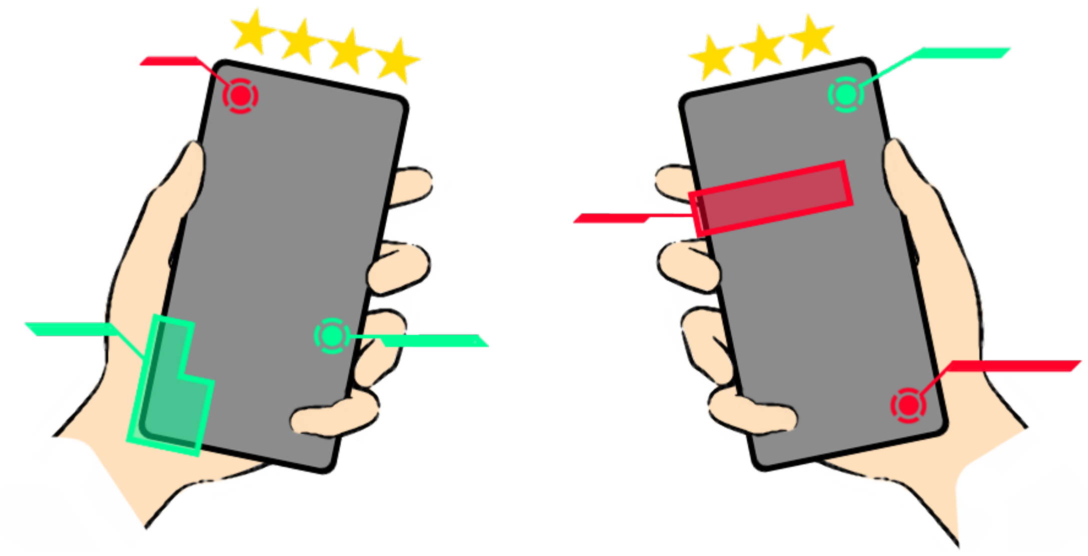


Figure 3: Concept for an AR comparison tool.

technical data should be needed to make a good choice, so that recommendations are welcomed by the user and it feels worthwhile to utilize comparison aids.

4 Conclusions and Further Research

In this paper we have discussed the possibilities of using AR in conjunction to RS. A previous approach is firstly discussed, where recommendations are offered by an embodied virtual advisor, providing as well product usage explanation via 3D augmentations and exploring multimodal interaction. After the findings of this first attempt, a new research direction is presented, focusing on the feasibility of creating a visual comparison aid for physical products, its possible benefits when combined with RS and the new challenges that come along with them. In the near future, we aim to define a number of interaction and recommendation models that could work well with our approach, put these ideas into practice by creating various prototypes and conduct the corresponding user studies to evaluate them.

References

Alvarez, G. & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Andrienko, G. L. & Andrienko, N. V. (1999). Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, *13*(4), 355–374.

Chatzopoulos, D., Bermejo, C., Huang, Z., & Hui, P. (2017). Mobile Augmented Reality Survey: From Where We Are to Where We Go. *IEEE Access*, *PP*(99), 1–1.

Gentner, D. & Medina, J. (1997). Comparison and the Development of Cognition and Language. *Cognitive Studies*, *4*(1), 1_112–1_149.

Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, *10*(4), 289–309.

Gutiérrez, F., Cardoso, B. D. L. R. P., & Verbert, K. (2017). Phara: A personal health augmented reality assistant to support decision-making at grocery stores. In *Healthrecsys@recsys*.

Hopp, T. & Gangadharbatla, H. (2016). Novelty Effects in Augmented Reality Advertising Environments: The Influence of Exposure Time and Self-Efficacy. *Journal of Current Issues & Research in Advertising*, *37*(2), 113–130.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender Systems Handbook*. Springer US.

Sinha, R. & Swearingen, K. (2002). The role of transparency in recommender systems. In *CHI '02 extended abstracts on human factors in computing systems - CHI '02*. ACM Press.

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the conference on human factors in computing systems - CHI '03*. ACM Press.

Tintarev, N. & Masthoff, J. (2007). A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*. IEEE.

Tominski, C., Forsell, C., & Johansson, J. (2012). Interaction support for visual comparison inspired by natural behavior. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2719–2728.

Zhang, J., Jones, N., & Pu, P. (2008). A visual interface for critiquing-based recommender systems. In *Proceedings of the 9th ACM conference on electronic commerce - EC '08*. ACM Press.

Zhu, W. & Owen, C. B. (2008). Design of the promopad: An automated augmented-reality shopping assistant. *JOEUC*, *20*, 41–56.

Paper V. Augmented-Reality-Enhanced Product Comparison in Physical Retailing

Reproduced with permission from the Association for Computing Machinery: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Augmented-Reality-Enhanced Product Comparison in Physical Retailing*. In *Proceedings of Mensch und Computer 2019*, MuC'19, pp. 55–65.

https://doi.org/10.1145/3340764.3340800

Augmented-Reality-Enhanced Product Comparison in Physical Retailing

Jesús Omar Álvarez Márquez
Jürgen Ziegler
jesus.alvarez-marquez@uni-due.de
juergen.ziegler@uni-due.de
Universty of Duisburg-Essen
Duisburg, Germany

ABSTRACT

Augmented reality technology has experienced great improvement in recent years and it has been successfully applied to industry and entertainment settings. However, its application in everyday contexts such as shopping is still very limited. One of the requirements to seamlessly incorporate augmented reality into everyday tasks is to find intuitive, natural methods to make use of it. Due to the inherent capabilities of augmented reality to work as a visual aid to explore and extend the knowledge a user has of the surroundings, this paper proposes the combination of AR technology and product advisors in a novel approach for product comparison. The user's awareness of the differences between multiple physically present objects is enhanced through virtual augmentations, supporting an intuitive way of comparing two or more products while shopping. To assess the validity of the concept, a prototype for an AR-based shopping assistant for comparing vacuum cleaners has been implemented and evaluated in a user study, testing different methods of visual comparison and interaction.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Information visualization**.

KEYWORDS

augmented reality, physical object comparison, comparison visualization, natural interaction

MuC '19, September 8–11, 2019, Hamburg, Germany

ACM ISBN 978-1-4503-7198-8/19/09…$15.00
https://doi.org/10.1145/3340764.3340800

ACM Reference Format:
Jesús Omar Álvarez Márquez and Jürgen Ziegler. 2019. Augmented-Reality-Enhanced Product Comparison in Physical Retailing. In *Mensch und Computer 2019 (MuC '19), September 8–11, 2019, Hamburg, Germany.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3340764.3340800

1 INTRODUCTION AND BACKGROUND

Nowadays, augmented reality (AR) technology is becoming more readily available and holds promise for a multitude of application areas. People's interest on knowing and using the technology has increased too and the number of companies willing to spend resources in adding AR solutions to their products and working processes keeps rising [24, 25]. However, despite the attractiveness of the new technology and its potential to engage consumers [6], the novelty of AR fades away rather quickly [14] and it is difficult to find reasons to use AR technology regularly instead of a more common (and probably convenient) method [15].

On a related note, brick and mortar stores are starting to enhance the shopping experience with the inclusion of computer technologies like smart carts [17], smart shelves [5], RFID sensors [26] or the adoption of the Internet of Things technology [7]. AR has the potential of adding a further quality by bringing the physical and virtual shopping experience together. Especially interesting is its possible application to convey information about physical products or even to work as a personal shop assistant, particularly when the products in question require technical knowledge or the support of an expert opinion to prevent a wrong buying choice. The adoption of AR-based shopping assistants promises to be a beneficial approach for both, retailers and consumers, making the former more competitive and the latter more aware of their buying decisions.

Finding intuitive, natural ways to display and transmit product information requires to take into account current research on customer behaviour. Studies assure that the expected behaviour of a client when in a physical store is to focus on product characteristics, in a decision-making process that requires comparing attributes of different products against each other or against personal preferences [20]. Comparing is one of the most basic cognitive activities and plays an important role in understanding, discovering and evaluating our surroundings [9]. Nonetheless, if the items to be compared are many, retaining their characteristics could impose a big constraint, emphasized by the limitations of short-term memory [3]. In this regard, it has been proven that using spatial superimposition via AR helps to alleviate the mental workload of retaining information by eliminating short-term memory demands [27].

Based on the aforementioned arguments, we present here our approach to physical object comparison through AR technology, where differences between items are highlighted to ease the limitations encountered during the comparison process, aiming to answer the following research questions:

RQ1 What are suitable visualization methods for product comparison?

RQ2 Which interaction style is more convenient for exploring product attributes in a comparative manner?

To validate our approach, a prototype for Microsoft HoloLens has been implemented and evaluated in two different studies. The first one tested two comparison visualizations (total vs relative differences) while the second one focused on interaction methods (tap-based vs head-gaze-based).

2 RELATED WORK

Supporting the comparison of complex data objects has been thoroughly explored from a wide number of perspectives. Gleicher et al. [11] survey on visual comparison research provides a large list of references and establishes a taxonomy of visual designs for comparison, dividing the comparative space in three different categories: juxtaposition (showing different objects separately), superposition (overlaying objects in the same space) and explicit encoding of relationships (computing the relation between objects and visualizing it). Also relevant is the work described in Tominski et al. [28], where a general interaction concept to support comparison tasks in visualization is developed, stressing the great importance of mimicking natural behaviour.

In relation to our interaction concept, head-based techniques have received the interest of the scientific community when applied to 3D environments, especially after being adopted as the standard selection method by popular AR and VR devices (as is the case of the HoloLens or the Oculus Rift). Early work on this area can be seen in Mine [21], where navigation and selection through head movement is included. More recently, Esteves et al. [8] studies the accuracy of head-based input (gaze), proposing a technique for augmented reality based on it and [19] provides an extensive comparison between different multimodal techniques for precision target selection in AR and investigates the combination of eye and head-based tracking.

AR usage in the shopping context has been proposed and studied too. Examples of AR being utilized to enhance the information a client has of the products are the Promopad [29] or the more recent systems presented in Gutiérrez et al. [12] (focussed on providing health-related information of individual items), Ahn et al. [1] (oriented to support product exploration) and Rashid et al. [23] (an approach for browsing physical product shelves).

To the best of our knowledge, visually expressing feature differences and/or similarities of two or more physical objects has not been studied in conjunction with AR yet. Nonetheless, studies dealing with the comparison of physical objects against their digital counterparts could be taken as a reference, as in Georgel et al. [10], where an approach for "discrepancy check" in construction sites is presented.

3 SUPPORTING THE COMPARISON OF PHYSICAL OBJECTS VIA AUGMENTED REALITY

We aim to find suitable interaction and visualization methods for exploring and comparing products in a physical store situation, where augmentations of the available products are provided through a head mounted display (HMD). These augmentations are used to emphasize differences between selected products, making them more noticeable and granting effortless access to information that must otherwise be taken from product flyers or manuals or provided by a human shop assistant. This approach should ease the comparison process, allowing the user to directly visualize the differences of two products regardless of their location within the shop or for which many attributes are available. Thus, it is also within the scope of this research to investigate how such comparison visualization will impact the decision making phase that takes place during the buying process.

Comparison Visualization Methods

Regular comparison means often rely on the use of tables listing product attributes on a side-by-side view, where the information is disconnected from the related physical products. This method may be convenient for online retailers, but it is less than optimal when it comes to brick and mortar stores where customers would need to go back and forth from the real product to its related characteristics. AR, on the other hand, allows for a more direct access to product information, where attributes are shown anchored to the physical object they belong to; then, when the comparison occurs, values from the other compared objects are displayed next to the attributes of the current one, keeping the information attached to the product. In this way, all the information is available even if only one of the products is within the field of view of the customer. Nonetheless, technical products usually have many attributes and filtering techniques may be necessary to prevent overcrowding the display with them, such as the use of attribute categories to only show those chosen by the customer.

Through AR it is possible not only to show superimposed product attributes, but also to link the displayed information to the related parts of the product. For instance, when targeting a vacuum cleaner as a product of interest, it is feasible to visually locate where the dust container is placed. Moreover, numerical data like the dust container's capacity could be enforced by showing a 1:1 3D model of it aligned against the real product, further clarifying the meaning of the number and even directly comparing it, side by side or over the same space, against the capacity of a previously selected item. A main advantage of using AR instead of a traditional approach to product comparison is precisely this: being capable of comparing more than the regular textual or numeric data, but also offering a visualization of what they represent, their exact location, measurements or usage, aspects apt to be compared per se and hardly representable via a different medium.

As mentioned in the related work section, the scientific community has already shown interest regarding the comparison of text and numbers. In our case we mainly make use of a side by side view of these types of values, exploring two different visualization methods (based on the "juxtaposition" and "explicit encodings" categories proposed by Gleicher et al. [11]):

Absolute values Values are presented side by side as they are. Values corresponding to the current item are emphasized, while the values of other selected products are shown next to them, without being modified.

Relative values Values are presented side by side modified depending on how much they differ. Values of the current item are shown unmodified, while the modified values of previously selected items are displayed next to them (e.g. if the current item has a price tag of 50€ and it is being compared against another one that costs 56€, the displayed value will be 50€ and +6€, respectively). When dealing with non numeric values, the performance of the previously selected item regarding this specific attribute is estimated and then an arrow up (better than current item) or down (worse) is used instead (Figure 1).

Interaction Techniques

Performing actions in the digital world trough a HMD may still not feel completely intuitive for a majority of users. People have grown comfortable using traditional user interfaces and conventional interaction mechanisms, thus the importance of making the transition to this new reality as smooth as possible (preventing situations where users could feel lost or incapable to continue without external guidance). However, the use of a HMD allows for a different type of interaction consisting on the activation of virtual elements based on where the user's head is aiming to. The question arises about what method would perform better when exploring the attributes of different products via AR: to make use of those mechanisms users are aware of and utilize with regularity for interacting with digital elements (like the click action) or to employ a technique perhaps more fitting to the nature of a HMD when dealing with real objects (like looking at something). In this regard, our research evaluates two different interaction methods:

Explicit activation This type of interaction makes reference to how users communicate with digital elements by tapping (clicking) on them. In our case, that means that the user will be able to select products, explore and access the different parts of the UI by tapping on holograms or detected real objects.

Implicit activation In a similar fashion to how people show interest for the things around them or inspect the characteristics of a certain object, interaction is carried out through head gaze by pointing at the different UI elements during a dwell time of 0.75 seconds. The value was chosen to be within the limitations regarding application response times, where less than 0.1 seconds feels like the system is reacting instantaneously and more than 1 second may interrupt the user's flow of thought [22].

Figure 1: Absolute vs relative comparison for non-numeric values (for the attribute "filter type").

Prototype

To put this concepts into practice, a prototype AR-based shopping assistant has been implemented for Microsoft's HoloLens platform. A HMD approach was chosen over a different platform because it allows for more interesting interaction possibilities, also leaving the user's hands free to perform a direct inspection of the product. For evaluating purposes the prototype has been conceived to support the comparison of physical vacuum cleaners, although the approach could be easily transposed to different domains. Some of its features are:

Product information visualization There is a distinction between attributes that are linked to a certain part of a product and those that are related to the product in a more general sense. Attributes not related to a specific part are shown floating around it, while the ones referencing parts are "attached" to them. They are organized by the categories "comfort", "performance", "versatility", "maintenance", "filtration" and "accessories" (Figure 2). The categories view displays how well the vacuum cleaner performs on each one of them (using a 1-5 scale), working as a summary of its features. The score given to each category is calculated using a number of custom rules based on how specialized websites assess the quality of the attributes within the category. By selecting one of the categories, the user can access the specific attributes that have an impact on its score (Figure 3).

Attribute explanations In many cases, buying decisions are aggravated when the products have attributes for which expert knowledge is required. To assist costumers in that regard, three different mechanisms have been included, activated when a user selects a single attribute:

- Some attributes act as categories themselves, and selecting them will disclose more specific features. For instance, on the regular view the user can only see (and compare) whether the product includes or not a battery; by selecting the battery attribute other related features will show up, such us battery type, capacity or charging time, allowing for their comparison too (Figure 4).
- A button that displays further information appears, providing a deeper insight about the meaning of the attribute (Figure 5).
- If the attribute is linked to a physical part of the product, such part is highlighted too, showing its shape and location.

Product selection The system knows the morphology of each available product and, once recognized (via the use of markers and the Vuforia SDK), the user can select them by directly tapping on the real object. The system allows the selection of up to three different vacuum cleaners at the same time. Selecting a product will highlight it with a unique colour and include it in the comparison view.

Product comparison When two or three vacuum cleaners are selected at the same time, the comparison view is activated (Figure 3). It comprises the following elements:

- Side by side values: Attributes of the selected products are placed side by side on every chosen vacuum cleaner, distinguishing them by their highlight colours. Values shown will depend on what type of comparison visualization is being used (absolute or relative values).
- Best values: each attribute of the vacuum cleaner which is currently within the field of view is evaluated following the same set of rules used for scoring categories, determining whether it has the best value among the chosen products (highlighted in green) or not (in which case it will appear in red). Besides,

Figure 2: Comparison view of a product in the category section. Current product's ratings are shown in orange, while the ones of a previously selected one appear in blue.

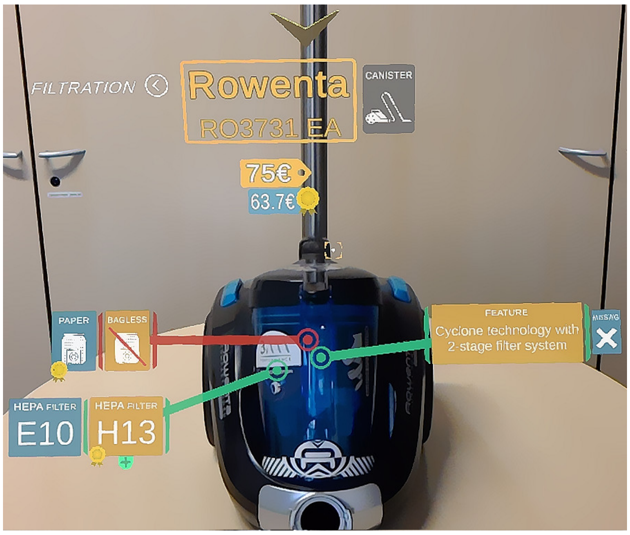


Figure 3: Comparison view of a product within the "filtration" category. Attributes linked to physical parts are attached to them through a line with a circled end.

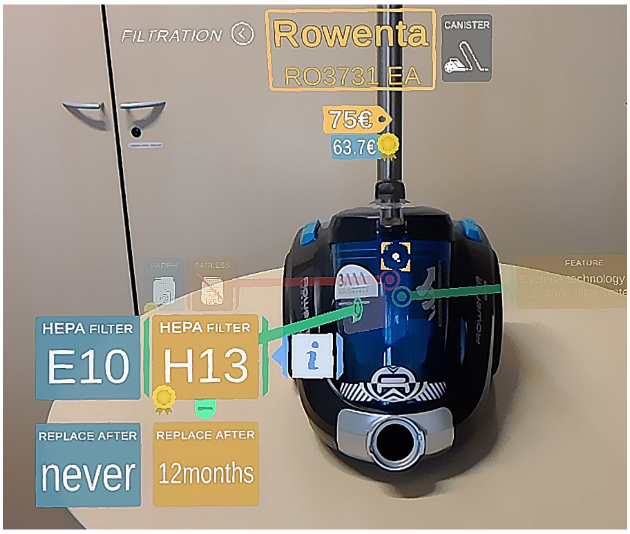


Figure 4: Selecting an attribute opens access to other related attributes (duration of the filter, in this case).

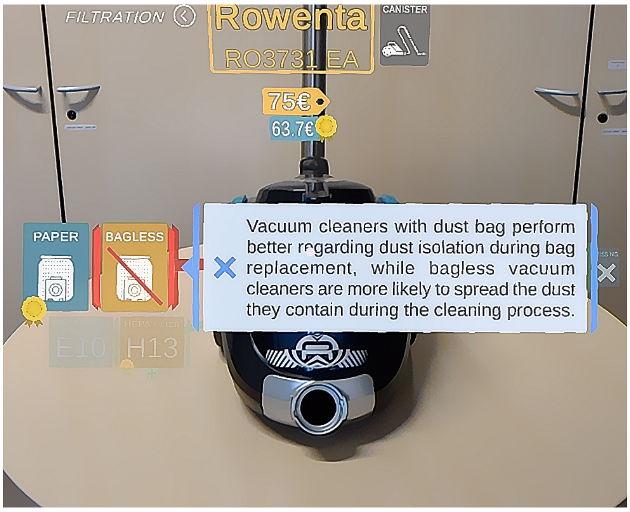


Figure 5: Information relative to the selected attribute.

the best value of a certain attribute among all the selected products appears with a golden badge attached to it.

Interaction The system implements both activation methods previously discussed. Category exploration and attribute selection can both be triggered via implicit or explicit activation, depending on which method is enabled. For vacuum cleaner selection only explicit activation is possible; the use of implicit activation for this purpose was discarded after some testing, which showed that trying to avoid undesired selections required great part of the user's attention.

Table 1: Characteristics of the available products

Feature	VC1	VC2	VC3
Type	handheld and upright	canister	wet-dry
Size	small	medium	big
Weight	light	medium	heavy
Capacity	low	medium	big
Suction power	low	medium	high
Battery	yes	no	no
Action radius	unlimited	large	short
Filter	bad	good	very good
Bag	no	no	yes
Accessories	many	few	many
Price	expensive	average	cheaper
Other	lights	cyclone tech.	blowing

4 EVALUATION

To evaluate our approach, two different studies have been conducted, each of them addressing one of our research questions.

First study: Absolute vs Relative Values

The main objectives of this study were to determine the usability of the approach and to analyse possible significant differences between the two described comparison visualization methods for value representation.

Settings and experimental tasks. Three different vacuum cleaner models were used. They cover different areas of usage but are similar enough so that they can be compared. Their main characteristics are shown in Table 1.

For the study, a floating canvas with instructions was added to the prototype, guiding users through the experiment. On it, three scenarios describing the client's specific needs were presented, asking participants to explore the attributes of the physical vacuum cleaners to find which one would cover the requirements. The given scenarios were:

- Small flat in the city. The tenant has a hairy dog. It also includes the attic, very dark due to the lack of windows.
- Family house with three floors. It has a large backyard with two big trees and lots of dead leaves during autumn. The family's car is usually full of dirt too.
- An elderly person with back problems, living in an old flat where space is scarce. She has a Persian cat and likes gardening in the balcony.

Notice that there was no completely right answer for the available vacuum cleaners in an attempt to increase item exploration. For instance, while the small, battery-powered vacuum cleaner with frontal lights could match most of the requirements for the first scenario, it has poor filtration ca-pabilities which would not be appropriate for a dog owner.

Three versions of the system were implemented, based on different visualization methods for comparing values:

(1) **No comparison enabled (NC)**: a version of the system where comparing items is not available, used as baseline. Participants are still able to visualize and explore product attributes individually.
(2) **Comparison through absolute values (AC)**: comparison is enabled, presenting unmodified values.
(3) **Comparison through relative values (RC)**: comparison is enabled, displaying absolute values for the vacuum cleaner that is within sight, but using relative values for the attributes of the items against this one is being compared.

Method. A total of 50 participants (38 female, average age of 21.16, σ 3.525) took part on the experiment. A between-subjects design was chosen, where only one of the implemented versions of the system was tested by a participant (16 tried NC, 17 AC and 17 RC). Individually, they were taught basic HoloLens usage and interaction possibilities offered by the prototype (for this study only explicit activation was available). After a couple of minutes for letting them get used to it and solve any possible questions they might have, they were told to follow the instructions given by the application and solve the three aforementioned scenarios, which were presented sequentially. After completing all of them, they were given a questionnaire covering aspects related to the ease-of-use of the system. It comprised SUS [4], AttrakDiff [13] and system-specific items measuring the constructs *content quality*, *usefulness* and *future usage intention*. A question directly addressing the preferred kind of visualization for different data types was added too, distinguishing between small numeric, big numeric and non-numeric values.

Results. Regarding the AttrakDiff items, the results can be seen in Table 2 and Figures 6 and 7. All three versions performed very well, falling into the "desired" category, although the one with no comparison received slightly worse results. Similar were the scores reported by the SUS items: 82.81 for the NC version, 83.97 for AC and 85.29 in the case of RC, which qualifies them as "excellent".

Concerning the question about what kind of visualization was preferred depending on the data type, 94% and 96% of the participants chose absolute values for small and big numbers, respectively. For the non-numeric values, 40% selected absolute values, while 60% of them liked the relative comparison more. It has to be noted that the relative comparison visualization in the case of non-numeric values highly differs from the one used for numerals, as previously shown in Figure 1.

Table 2: AttrakDiff's pragmatic (PQ) and hedonic qualities (HQ), along with their respective confidence and the system attractiveness (ATT)*

Version	PQ	Conf.	HQ	Conf.	ATT
No comparison	1.47	0.49	1.44	0.45	1.69
Absolute values	1.75	0.27	1.66	0.41	1.74
Relative values	1.65	0.48	1.74	0.47	1.91

* values provided by http://www.attrakdiff.de

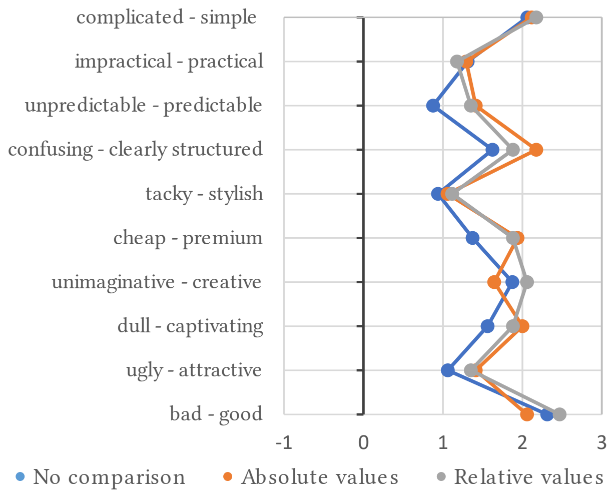


Figure 6: Mean values for AttrakDiff's word pairs in each system version. It uses a -3 to 3 scale, shortened to ease its visualization.

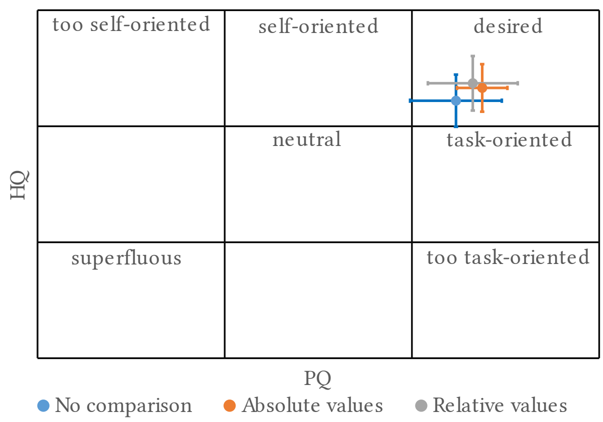


Figure 7: AttrakDiff's pragmatic and hedonic qualities

Table 3: System-specific items

Question	No Comparison mean	No Comparison σ	Absolute Values mean	Absolute Values σ	Relative Values mean	Relative Values σ
I found the system helpful for deciding for a fitting product.	4.31	1.01	4.35	0.70	4.47	0.62
I found the system helpful for discovering and understanding the product's attributes	4.38	0.61	4.47	0.51	4.65	0.49
The system saved me time in understanding differences between products.	3.38	0.95	3.71	1.10	4.24	1.03
If both options were available, I would prefer to use the AR system instead of a conventional product comparison table	3.19	1.27	3.47	1.23	3.59	1.33

Table 3 shows a list of items covering system-specific aspects not observed by the SUS nor the AttrakDiff questionnaires. A One-way between groups ANOVA analysis was performed over all the items, but there were no significant results.

Discussion. Although there were no significant differences between comparison methods, both of them performed better than the system with no comparison capabilities. It is interesting that despite relative comparison performing slightly better in terms of SUS score and being perceived as more "creative", "good" and "helpful for understanding differences between products", participants would mostly prefer to use absolute comparison instead. It thus seems promising to implement a mixed comparison visualization method that takes into account the type of data to be compared, considering that more than half of the participants chose relative over absolute comparison for non-numeric values, but there was consensus about only using absolute comparison when dealing with numbers. Also, it has to be noted the existence of different, unexplored ways to show relative differences that may have performed better (e.g. the use of percentages) and for which further research is needed. Finally, the novelty of the technology has to be taken into account too, being more than possible that part of the achieved high scores in both SUS and AttrakDiff rely on the fact of using AR.

Second study: Explicit vs Implicit Activation

A second study was performed aiming to explore the possible implications of using implicit vs explicit activation.

Settings and experimental tasks. Mirroring the first study, the same three vacuum cleaners have been used. Equally, participants had to complete the same tasks than before. In relation to our second research question, two new system variations where added based on different input methods:

(1) **Explicit activation (EA)**: activating an attribute and accessing to its information requires a "tap" action.
(2) **Implicit activation (IA)**: using the head to gaze at an attribute activates it, bringing it closer to the user after the dwell time (0.75s) has passed. For accessing its information, the user must aim directly into the pertinent icon. In both cases the pointer's shape will change to display a loading bar, making the user aware of the remaining dwell time before the attribute is selected.

Method. 29 participants took part in the study (17 female, average age of 22.9, σ 3.31). A mixed design was chosen for the experiment, using a within subjects approach for visualization methods and an between-subjects design for activation techniques. To compensate for the impact of the order in which the comparison versions were tested, the conditions where appropriately counterbalanced. By the end, 15 subjects tested EA and 14 IA.

As in the former study, participants where firstly instructed in the usage of the HoloLens. After clarifying any questions, the first task was presented to them and accomplished straight away by using one of the three comparison versions. After its completion, subjects were asked to fill in a short questionnaire. The same procedure was repeated two more times, one per comparison method. Through the whole process, only one activation technique was available.

The questionnaire was composed by the short version of the UEQ proposed in [2] and a set of system-specific items measuring *content quality*, *usefulness* and *future usage intention.*

Results. The results reported by the UEQ questionnaire are presented in Figures 8 and 9. Scores obtained by comparison method confirm the results reported in the previous study, showing values very similar for the three different versions, maintaining the no-comparison one in the worst position. When considering the different methods of activation, IA is perceived as more "creative", "motivating" and "valuable" than EA by a noteworthy difference, while displaying almost equivalent values for the rest of items.

Questionnaire dimensions: A mixed ANOVA repeated measures analysis was performed over the scores obtained for the dimensions measured by the UEQ and the system specific questions (Table 4). The test reported a statistically significant interaction effect between activation and comparison methods for the dimension "hedonic quality" ($F(2, 54) = 5.098, p < .01, \eta_p^2 = .16$). Searching for possible simple main effects, a further multivariate ANOVA test of the between-subjects factor indicated that there is a significant difference in hedonic quality scores between activation methods when using relative comparison ($p < .005$), for which IA obtained better results. Likewise, possible simple main effects of the within-subjects factor were analysed via a one-way repeated measures ANOVA for each activation method separately. For the group using EA, there was a statistically significant effect of the comparison method on hedonic quality scores ($F(2, 28) = 4.485, p < .05, \eta_p^2 = .25$) and performing a Bonferroni test confirmed that there was a significant difference between not using comparison at all and using relative comparison in terms of hedonic quality ($p < .05$), but only when EA is enabled. It has to be noted that, among all the possible combinations, the use of explicit activation with relative values visualization received the lowest score for this dimension. Possible main effects concerning the rest of the dimensions not affected by the interaction between factors were analysed too, but no significant differences were found for any of the independent variables.

Empirical data: Table 5 contains the log data collected during the experiment. A mixed ANOVA repeated measures analysis indicated that there was no statistically significant interaction between activation and comparison methods for any of the measured dependent variables, reason for which the results are omitted here. Equally, no significant main effects were found for comparison methods. Nonetheless, there were significant main effects for activation type, meaning that intervention groups differed significantly, regarding the times an attribute was gazed ($F(1, 27) = 6.49, p < .05, \eta_p^2 = .2$) and tapped ($F(1, 27) = 18.3, p < .001, \eta_p^2 = .41$) as well as how often participants used the help option ($F(1, 27) = 18.42, p < .001, \eta_p^2 = .41$). A closer look to the collected data (Table 5) shows that the times per task were slightly higher for the system without comparison means in both activation techniques, whereas IA shows greater values overall and reaches the highest time per task of all the possible combinations when in conjunction with NC (40% longer than any other time). With EA enabled, participants tended to gaze more into attributes by a great extent. In contrast, during the sessions with IA, subjects selected attributes up to three times more often. Participants' need for visually switching between vacuum cleaners remained fairly consistent in both activation alternatives, although slightly lower values were

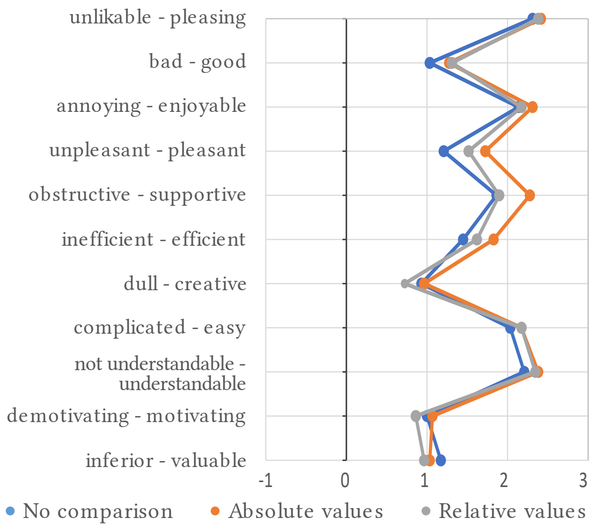


Figure 8: UEQ's items by comparison method. It uses a -3 to 3 scale, shortened to ease its visualization.

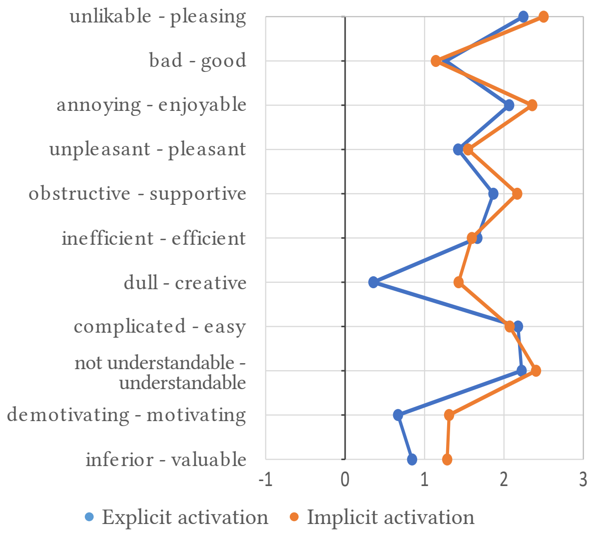


Figure 9: UEQ's items by activation method. It uses a -3 to 3 scale, shortened to ease its visualization.

collected when using RC. On the other hand, the usage of the "information" feature increased greatly with IA and, within this technique, in systems with comparison enabled it was used twice as often than for the no comparison one. Lastly, not all participants made use of the comparison view, preferring to explore each vacuum cleaner and their attributes individually. This is represented by the low percentages of users triggering the comparison, especially when using relative value visualization in conjunction to EA. Comparing three vacuum cleaners at once was a seldom choice in general, showing a modest increment with IA.

Table 4: Mean values for the dimensions measured by the questionnaire. The three first ones are extracted from the UEQ (-3 to 3 scale), while "USE" and "PRE" are system specific (Likert 5 point scale), addressing how useful the prototypes were for exploring/finding products and the participant's preference of usage of the system over a traditional mean, respectively.

	Implicit Activation								Explicit Activation							
	No Comparison		Absolute Comparison		Relative Comparison		Total		No Comparison		Absolute Comparison		Relative Comparison		Total	
	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ	mean	σ
ATT	1.71	0.73	1.98	0.53	1.96	0.61	1.88	0.13	1.63	0.73	1.88	0.62	1.73	0.71	1.75	0.15
PQ	1.96	0.67	2.10	0.57	2.10	0.65	2.06	0.14	1.81	0.79	2.21	0.53	1.91	0.55	1.98	0.12
HQ	1.33	0.70	1.26	0.64	1.42	0.59	1.34	0.15	0.75	0.88	0.80	0.79	0.31	1.13	0.62	0.22
USE†	3.80	0.74	4.04	0.67	4.14	0.93	4.00	0.19	4.04	0.64	4.22	0.55	4.06	0.52	4.11	0.13
PRE†	3.82	0.74	3.96	0.86	3.96	0.81	3.91	0.18	3.91	0.87	3.90	0.76	3.89	0.68	3.90	0.18

† Likert 5 point scale

Table 5: Mean values for the measured empirical factors: required time per task in minutes, number of times a user gazed and tapped (selected) an attribute, how often participants switched their attention from one vacuum cleaner to another, in how many occasions an attribute's information was inspected and the percentage of users who made intentional use of the comparison view of two and three vacuum cleaners (that is, explored items with that view at least during 1 minute).

Version	Time per task	Gaze on attribute	Select attribute	Switch products	Check information	Compare 2 products	Compare 3 products
EA - NC	04:31	152.00	11.47	61.33	3.07	-	-
EA - RC	03:55	132.87	9.13	50.47	0.47	40%	20%
EA - AC	04:28	154.20	11.73	58.87	1.87	80%	20%
IA - NC	07:02	96.00	24.50	62.70	6.64	-	-
IA - RC	04:45	95.85	26.50	60.43	12.21	64%	36%
IA - AC	05:01	90.86	30.86	66.14	11.50	71%	29%

Discussion. The possible benefits of using a comparison technique against not using one are still unclear. In addition to their very similar UEQ results, users appear to visually switch between items with the same frequency no matter whether the comparison is enabled or not, even though the assumption would be to obtain lower values when it is enabled. Although all the pertinent attributes of the item which is out-of-sight are displayed next to the ones belonging to the product the user is currently examining, participants still have the need to visualize the out-of-sight one. Something to consider in this regard is that the products where placed near to each other during the study, which may not be the case in a real world situation. Perhaps the advantages of using the comparison feature would have been more obvious with the added effort of having to walk to a distant product. Testing the system with a set-up closer to a real store may provide further insight about this issue. Despite this, times per task seem to be lower when using a comparison view, suggesting a faster acquisition of information.

Coming back to our first research question, it has to be noted that the comparison view was activated less often when using relative values, probably in association with the results of the previous study where most of the participants chose the absolute values visualization over the relative one. The disinclination to use the relative-values view may be connected to participants having to mentally calculate quantities when applied to numerical data, issue that could be solved by the use of percentages instead of raw numbers. Interestingly, this only seems to affect to the perceived hedonic quality of the system when combined with explicit activation. In general, absolute values visualization seems to be the preferred method for most situations, although the results regarding comparison methods leave open questions and further research is required in this direction, especially considering alternative ways of presenting relative differences and the reasons behind product-switching not being lessened by the usage of the comparison view.

With respect to interaction techniques, the inclusion of implicit activation seems to have a significant effect on the perceived hedonic quality of the system, also reflected in Figure 9, where the three scales in which IA clearly outperforms EA are the ones from which the hedonic quality is calculated. Hence, IA was perceived as more novel or stimulating in general but equally functional than EA (pragmatic quality), probably because IA is a kind of interaction the participants had not used before.

Subjects gazing into attributes significantly less often with IA could be explained by them trying to avoid undesired selections, which would also mean that a more careful navigation was taking place. This works as well as a possible justification for spending more time per task with IA than when using EA in general. It would explain too the very long time per task obtained for the use of implicit activation in a system with no comparison view, specially considering what has been mentioned earlier about the system without comparison means being less effective in terms of information acquisition. Accidentally selecting items is a well studied issue, known as the "Midas Touch" problem [16]. Giving the users awareness about when an attribute is on the course of being selected (like changing the pointer's shape into a timer) and choosing a proper dwell time are helpful methods to mitigate unwanted attribute selections, both of them implemented in the system. Nonetheless, the effects of such issue should still be considered when explaining the higher values for attribute selections during IA. Less probable are the negative effects of the Midas Touch affecting the number of times users accessed the extra information of an attribute: it is a two-step process where first an attribute has to be selected and then the information icon must be "tapped", which is a course of action less susceptible to errors. Given the significant difference between activation methods regarding information checks (also proportionally speaking when compared against attribute selections) it is relatively safe to assume that using IA influenced attribute exploration both in a negative and positive manner, requiring a more careful navigation than EA but at the same time encouraging a more active inspection of the attributes.

The question arises whether the improved hedonic quality and higher attribute inspection when using implicit activation compensate for a harder navigation. As presented in Karapanos et al. [18], the importance of the hedonic quality in terms of the incorporation of a product in daily routines display a sharp decrease after some time has passed, contrary to what happens to the pragmatic quality. This means that the significant difference between hedonic quality scores of both activation methods may not be so after a prolonged usage experience. It is to be expected that participants would have preferred explicit activation in the long run, because it is a less complicated and more direct interaction method and no other significant differences were found. However, it would be interesting to further explore the aspects of implicit activation that encouraged the users to look into the information of the attributes more often and how to take advantage of them when using explicit interaction.

5 CONCLUSIONS AND FURTHER RESEARCH

An approach to product comparison supported by augmented reality has been presented, aiming to enhance the intuitive action of visually comparing two different products when buying in a physical store situation. Augmentations of a product's attributes are shown next to it, also giving further insight about their meaning and highlighting related product parts. Customers can freely select various products, activating a comparison view when two or three are chosen at the same time. The comparison view places the individual attributes of the selected products side by side (emphasizing their differences) and alleviates the effort of remembering them, supporting the decision making process of choosing which one to buy.

A prototype AR-based shopping assistant has been implemented for Microsoft's HoloLens and evaluated in two different user studies. There was a very positive overall outcome in terms of user experience and satisfaction. Results suggest that the inclusion of a comparison feature has a low impact in that regard, although they also indicate a quicker information acquisition when comparison is enabled. When comparing the values of an attribute, their absolute (unmodified) presentation was generally preferred. The implicit selection of attributes through head gaze obtained better results in terms of hedonic quality and attribute examination, but requiring a more careful navigation. Previous studies indicate that in the long run explicit activation may be preferred.

Further research includes looking into the reasons behind the seemingly counter-intuitive results regarding the usage of comparison methods and studying new visualization ways for displaying relative differences. It is also important to identify which aspects of the implicit activation boost the inspection of attributes and find possible ways to apply them to explicit interaction. Additional work is planned for including a recommender system that works on top of the already implemented prototype, giving recommendations not only in the shape of fitting products, but also concerning attributes to be shown (thus, minimizing the need of exploring attributes through the use of categories). It is also in our scope to include a digital catalogue of products, allowing the comparison between physical and non-physical items.

REFERENCES

[1] Junho Ahn, James Williamson, Mike Gartrell, Richard Han, Qin Lv, and Shivakant Mishra. 2015. Supporting Healthy Grocery Shopping via Mobile Augmented Reality. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 1s (Oct. 2015), 16:1–16:24. https://doi.org/10.1145/2808207

[2] Catherine Alberola, Götz Walter, and Henning Brau. 2018. Creation of a Short Version of the User Experience Questionnaire UEQ. *i-com* 17, 1 (apr 2018), 57–64. https://doi.org/10.1515/icom-2017-0032

[3] G.A. Alvarez and P. Cavanagh. 2004. The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science* 15, 2 (feb 2004), 106–111.

[4] John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[5] Steve Brown and David Roth. 2015. The Future of Retail: Shopping and the Smart Shelf. *Intel* (2015). https://www.intel.com/content/www/uk/en/retail/digital-retail-futurecasting-report.html

[6] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. 2017. Mobile Augmented Reality Survey: From Where We Are to Where We Go. *IEEE Access* PP, 99 (2017), 1–1.

[7] Rong Chen, Li Peng, and Yi Qin. 2010. Supermarket shopping guide system based on Internet of things. In *IET International Conference on Wireless Sensor Network 2010 (IET-WSN 2010)*. IET. https://doi.org/10.1049/cp.2010.1020

[8] Augusto Esteves, David Verweij, Liza Suraiya, Rasel Islam, Youryang Lee, and Ian Oakley. 2017. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 167–178. https://doi.org/10.1145/3126594.3126616 event-place: Québec City, QC, Canada.

[9] Dedre Gentner and José Medina. 1997. Comparison and the Development of Cognition and Language. *Cognitive Studies* 4, 1 (March 1997), 1_112–1_149.

[10] Pierre Georgel, Pierre Schroeder, Selim Benhimane, Stefan Hinterstoisser, Mirko Appel, and Nassir Navab. 2007. An Industrial Augmented Reality Solution For Discrepancy Check. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)*. IEEE Computer Society, Washington, DC, USA, 1–4. https://doi.org/10.1109/ISMAR.2007.4538834

[11] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (sep 2011), 289–309.

[12] Francisco Gutiérrez, Katrien Verbert, and Nyi Nyi Htun. 2018. PHARA: An Augmented Reality Grocery Store Assistant. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '18)*. ACM, New York, NY, USA, 339–345. https://doi.org/10.1145/3236112.3236161

[13] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003: Interaktion in Bewegung*, Gerd Szwillus and Jürgen Ziegler (Eds.). Vieweg+Teubner Verlag, Wiesbaden, 187–196. https://doi.org/10.1007/978-3-322-80058-9_19

[14] Toby Hopp and Harsha Gangadharbatla. 2016. Novelty Effects in Augmented Reality Advertising Environments: The Influence of Exposure Time and Self-Efficacy. *Journal of Current Issues & Research in Advertising* 37, 2 (July 2016), 113–130.

[15] Tseng-Lung Huang and Shuling Liao. 2015. A model of acceptance of augmented-reality interactive technology: the moderating role of cognitive innovativeness. *Electronic Commerce Research* 15, 2 (June 2015), 269–295.

[16] R Jakob. 1998. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *Readings in intelligent user interfaces* (1998), 65–83.

[17] Gerrit Kahl, Lübomira Spassova, Johannes Schöning, Sven Gehring, and Antonio Krüger. 2011. IRL SmartCart - a User-adaptive Context-aware Interface for Shopping Assistance. In *Proceedings of the 16th International Conference on Intelligent User Interfaces (IUI '11)*. ACM, New York, NY, USA, 359–362. https://doi.org/10.1145/1943403.1943465 event-place: Palo Alto, CA, USA.

[18] Evangelos Karapanos, John Zimmerman, Jodi Forlizzi, and Jean-Bernard Martens. 2009. User experience over time. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. ACM Press. https://doi.org/10.1145/1518701.1518814

[19] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 81:1–81:14. https://doi.org/10.1145/3173574.3173655 event-place: Montreal QC, Canada.

[20] Kelvin J. Lancaster. 1966. A New Approach to Consumer Theory. *Journal of Political Economy* 74, 2 (April 1966), 132–157.

[21] Mark R. Mine. 1995. *Virtual Environment Interaction Techniques*. Technical Report. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[22] Jakob Nielsen. 1994. *Usability engineering*. Elsevier.

[23] Zulqarnain Rashid, Rafael Pous, Joan Melià-Seguí, and Marc Morenza-Cinos. 2014. Mobile Augmented Reality for Browsing Physical Spaces. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 155–158. https://doi.org/10.1145/2638728.2638796 event-place: Seattle, Washington.

[24] Alexandra Rese, Daniel Baier, Andreas Geyer-Schulz, and Stefanie Schreiber. 2017. How augmented reality apps are accepted by consumers: A comparative analysis using scales and opinions. *Technological Forecasting and Social Change* 124 (Nov. 2017), 306–319. https://doi.org/10.1016/j.techfore.2016.10.010

[25] Lucas Roitman, Jeff Shrager, and Terry Winograd. 2017. A Comparative Analysis of Augmented Reality Technologies and their Marketability in the Consumer Electronics Segment. *Journal of Biosensors & Bioelectronics* 08, 01 (2017). https://doi.org/10.4172/2155-6210.1000236

[26] Michael Schneider. 2003. A Smart Shopping Assistant utilising Adaptive Plan Recognition. *ABIS 2003 – 11. GI-Workshop "Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen"* (2003). http://dl.gi.de/handle/20.500.12116/5123

[27] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the conference on Human factors in computing systems - CHI '03*. ACM Press.

[28] C. Tominski, C. Forsell, and J. Johansson. 2012. Interaction Support for Visual Comparison Inspired by Natural Behavior. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (dec 2012), 2719–2728.

[29] Wei Zhu and Charles B. Owen. 2008. Design of the PromoPad: An Automated Augmented-Reality Shopping Assistant. *JOEUC* 20 (2008), 41–56.

Paper VI. In-Store Augmented Reality-Enabled Product Comparison and Recommendation

Reproduced with permission from the Association for Computing Machinery: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *In-Store Augmented Reality-Enabled Product Comparison and Recommendation*. In *Fourteenth ACM Conference on Recommender Systems*, RecSys '20, pp. 180–189.

https://doi.org/10.1145/3383313.3412266

In-Store Augmented Reality-Enabled Product Comparison and Recommendation

Jesús Omar Álvarez Márquez
University of Duisburg-Essen
Duisburg, Germany
jesus.alvarez-marquez@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT

We present an approach combining the AR-based presentation of product attributes in a physical retail store with recommendations for items only available online. The system supports users' decision-making process by offering functions for comparing product features between items, both physical and online, and by providing recommendations based on selecting in-store products. The physical products may thus serve as anchors for forming the user's preferences, also offering a richer and more engaging experience when exploring the products hands-on. Both objective product attributes as well as the visual appearance of a physical product are employed for generating recommendations from the online space. In this way, the advantages of online and in-store shopping can be combined, creating novel multi-channel opportunities for businesses. An empirical evaluation showed that the comparison and recommendation functions were appreciated by users, and hinted some possible benefits of a hybrid physical-online shopping support system. Despite the limitations of the study, there is sufficient evidence to consider this a viable approach worth to be further explored.

CCS CONCEPTS

• **Human-centered computing** → *Human computer interaction (HCI)*; *Information visualization*; *Mixed / augmented reality*; • **Information systems** → *Recommender systems.*

KEYWORDS

augmented reality, recommender systems, preference construction, product comparison

ACM Reference Format:
Jesús Omar Álvarez Márquez and Jürgen Ziegler. 2020. In-Store Augmented Reality-Enabled Product Comparison and Recommendation. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3383313.3412266

RecSys '20, September 22–26, 2020, Virtual Event, Brazil

ACM ISBN 978-1-4503-7583-2/20/09…$15.00
https://doi.org/10.1145/3383313.3412266

1 INTRODUCTION

Even though the idea of augmenting the physical world via the inclusion of digital elements is not a new concept at all [3], the technology that allows for it to be usable on a daily basis has only recently become available on a wider basis. It is in the working environment where augmented reality (AR) has shown its most notable benefits, especially when it comes to improving productivity and quality of production or to providing training and assistance for complex tasks [30]. Entertainment and marketing are also areas where AR has seen more use due to the attractiveness of the new technology and its potential to engage consumers [10], to such an extent that the investment in AR solutions is expected to grow exponentially as the technology matures [29]. In marketing and retail, AR has been applied in various forms at mobile and local customer touch-points [7]. However, in many cases, the engaging effect of AR may rely considerably on its novelty which tends to decrease rather quickly [18] in favour of more conventional methods of interaction.

Combining AR with recommender technologies appears to be an avenue that offers the potential for creating both an engaging customer experience and pragmatic benefits in terms of search and decision support [1, 31, 46]. This combination, however, has thus far only been exploited in specific contexts, such as providing recommendations for mobile users [40, 45]. The application of AR-based recommender techniques in the physical setting of a retail store, in contrast has hardly been investigated yet [9].

Applying AR in a physical store offers various opportunities for supporting users in their decision-making process. The most obvious advantage is that AR can give users the option to explore the properties of a product in situ without the need to switch attention between product and additional information sources, such as leaflets or product websites. A further promising, yet unexplored, function relates to comparing the properties between two and more physically present products since AR can virtually combine product attributes and display them in the vicinity or as overlay of a product the user is looking at . Providing such functions can relieve the burden on the user's memory [2, 39] which can be considerable when comparing a larger number of seemingly similar products in a store . Seeing the properties of a product in direct spatial relation to the product and its parts may enable the user to criticize a product feature and ask for products with different feature values, which can then be recommended from the set of products available.

The recommender function becomes much more powerful, however, if recommendations from the vendor's online offerings can be included in the AR presentation. In this case, the presence of a physical product can help the user to construct his/her preferences more effectively, in particular when certain product features are best understood when it is possible to examine a product in its physical form. A hybrid approach combining physical and virtual products in an AR interface allows to use a small selection of physical products as reference points or anchors[42] for a larger online collection thus reducing space requirements and costs.

The approach is also in line with the ideas of omni-channel retailing [19, 44] where different communication channels cooperate for a more rewarding shopping experience (e.g. use internet to obtain product information when in a physical store). It can also be particularly helpful when dealing with products that require technical knowledge or the assistance of experts to prevent a wrong buying choice. Avoiding inventory limitations, information accessibility and ease of comparison are some of the features for which online shopping is usually preferred over physical stores [47]. With the hybrid approach described in this paper, these benefits can be brought into the physical environment.

In this paper, we present an AR-based shopping support system that combines product comparison and recommending methods for both physical and online products, significantly extending the ideas described in [49]. The concept revolves around the idea of letting users browse the digital product space by exploring the physical one. Its main features include the ability to display relevant attributes of physical products, to allow direct product comparison and to provide product recommendations. Physical products can be compared against each other and against digital ones. Furthermore, recommendations can be influenced by critiquing attributes of physical products.

Our research goals in the matter of bringing virtual recommendations into the physical shopping scene can be summarized in the following research questions:

RQ1 How effective are product recommendations provided through AR?
RQ2 How can the development of user's preferences be supported by AR-enhanced product displays?
RQ3 Does the presence of a physical product serve as a cognitive anchor for selecting among online products?

In accordance with the described approach, a prototype for a shopping support system for Microsoft HoloLens has been implemented and tested in a user study, for the purpose of answering our research questions and evaluating the usefulness of the system when it comes to the alleviation of the limitations of physical retailing.

In the following, related work is discussed, while successive sections describe in deeper detail our approach, the functionality of the prototype developed and the design and results of its evaluation.

2 RELATED WORK

2.1 Recommendations in physical retailing

Online stores typically include features like product comparison tools [23, 32], price trackers, customer reviews and ratings [27], detailed descriptions or product recommendations [37], all of them oriented to providing useful information that supports clients in their purchase decision and overcome the limitations of not having direct access to physical goods. On the other hand, obtaining such amount of information in a physical context depends mainly on the interaction with sales staff [17], although it is not always possible to have access to reliable information sources. The inclusion of recommender systems could be a solution to the information demands of clients in retail stores. As an example of this approach, Kourouthanassis et al. [25] presents a system that automatically creates a shopping list that is updated in real time when the user picks something at the store, while also offering product information and recommending promotions based upon previous buying behaviour or cross-selling associations. Another example would be APriori [45], a system for mobile devices that lets consumers receive product data, recommendations and user ratings directly at the point of sale. There are a few instances in the research field where AR-based recommendations have been used for providing in-store support. In that regard, a system that recommends healthy products is presented in Ahn et al. [1], where the authors also assess, among other aspects, the benefits of using AR for product search in retail stores; in Gutiérrez et al. [15] an AR shopping assistant is described, PHARA, that delivers health-related information, focusing their research on visualization layouts and their convenience for different AR platforms.

2.2 User preference models

Consumers do not always have well-defined preferences, but often tend to build them on the spot when making a decision is required [33]. The lack of preferences becomes an issue especially with digital catalogues where there is a great number of choices that have to be evaluated, possibly leading to choice overload [6]. Recommender systems play a key role in reducing the amount of information that consumers need to evaluate, while they also have the capacity to influence the client's preference-construction process [16]. There are several recommending-related factors that may have an impact on the creation of preferences, from the influence of numerical attributes [26] to the mere presence of recommendations [24]. In our research, however, it is the presence of physical products what could have an effect on the client's final decision, a factor that has not yet been considered due to how rarely physical and digital products are presented together. In this particular scenario, psychological effects such as priming[41] and anchoring[12] should be considered, where physical objects may influence a client's judgement on a perceptual or cognitive level. It is through the usage of physical products that the exploration of a larger set of digital ones is performed, thus supporting a progressive discovery of the product catalogue and the development of consumer preferences.

2.3 AR in retail stores

After years of confrontation between online and physical retailers, traditional companies have begun to understand that the future is digital, to the extent that most of them now offer online retailing channels that may work in parallel or in combination to the already existent physical ones. Depending on the level of integration among the available channels, retailers can be classified as multichannel, cross-channel or omni-channel [5, 20]. Omni-channel retailing stands for the greatest level of channel integration, where the boundaries between physical and virtual channels have disappeared to provide a seamless shopping experience to customers. The omni-channel approach is slowly taking the stage and replacing current multi-channel retailers [44] (for which each channel works independently), as demonstrated by the new functions that physical stores have gradually taken [13], such as pick-up points or showrooms.

AR in particular has gathered a lot of attention in the retailing context due to its capacity to increase consumer engagement and influence the purchase decision [31]. In-store AR applications have been made popular in the shape of virtual try-on (also called "magic-mirrors") that have gathered a lot of attention [4, 21, 22, 38]. An early example of AR being used in retail stores is The PromoPad [48] which is capable of providing context-aware information of products; Välkkynen et al. [43] developed an approach to visualize package contents before its opening; Rashid et al. [35] uses a combination of RFID with AR to browse physical product shelves; Acquia Labs created a demo [8] for an AR shopping assistant to showcase the possibilities of currently available technology, with features like the superimposition of useful information and customizable product search; Cruz et al. [11] created an AR mobile application for retail stores that detects where the user is located and provides guidance to the item that the user is looking for.

Despite the existence of previous research on AR-based in-store shopping assistants, their combination with product recommending features has rarely been explored so far. Additionally, psychological aspects that may play a role on user acceptance of the concept have been generally overlooked, as it could be the case of priming and anchoring effects.

3 AR-BASED RECOMMENDING AT THE POINT OF SALE

Most of the challenges related to the provision of recommendations can be summarized in three simple questions: what to show, when to show and how to show the information [36]. Rather than addressing these matters directly, we consider it necessary first to create a solid foundation on which recommendations can be built: in our case, it means to find basic activities that customers perform when buying in order to design solutions to support the recommendation aspect over them. That is, recommendations are built upon more elementary aspects that support the buying process in general, such as the access to external information, attribute explanations and comparison methods, which should be as integrated as possible within a normal buying behaviour.

Literature in the field of consumer behaviour point at the comparison of features of different buying options as one of the most common actions that clients of a physical store perform when making a purchase decision [28]. Comparing plays an important role in the client's decision-making process that precedes the selection of a product, which involves not only the comparison of the available items against each other but also against the customer's personal preferences. For this reason, we have taken the comparison of products as the foundation of our approach to AR-enabled shopping assistants. In it, clients can unveil the attributes of physical products, navigate them, learn about their meaning and compare them against the attributes of other products. On top of it, a recommender system has been designed to display products similar to the physical one at which the client is currently looking. These recommendations expand the limitations of the physical catalogue by enabling the selection of products that are not physically present at the store. The digital-product space is browsed by exploring the physical one, which opens room for interesting questions regarding the effects that real world items may have over the choice of digital ones when the former are used as a reference for the latter. Moreover, the recommender system allows user feedback by enabling attribute critiquing, which supports the creation of a mental preference model sustained over the examination of physically present products.

The outlined concept has been implemented into an application that runs under Microsoft's HoloLens and uses marker-based product detection (supported by the Vuforia Engine [1], which provides advanced computer vision functionality to recognize images and objects in AR applications). Although using a smartphone as AR enabler may be a more practical approach for present-day retailing, in this research a head mounted display (HMD) has been chosen instead, even if it means using a medium to which users are less accustomed and may be seen as adding complexity. The reasons behind this decision are: first, because HMD technology is becoming more relevant and it is likely to be more accessible in a near future; second, its nature makes it more interesting in retailing contexts, because it offers a more engaging experience (which is particularly relevant for advertising and promotion stands) and allows for a hands-free direct inspection of products without loosing sight of the augmentations; and third, research in this field is still immature and leaves more opportunities for future work.

The prototype here described is designed to work with physical vacuum cleaners (Fig. 1), but the concept itself could be applied to many different domains. More concrete aspects of the comparison and recommendation features and their implementation are explained over the following subsections.

3.1 Information access

In a normal set-up, clients of a physical store may only access product information by consulting flyers or asking a human sales person. In such scenario, consumers may face situations where the information provided is not sufficiently accurate or complete, or perhaps not enough personnel is available, or they do not have the required expertise to provide support. Human factor aside, even if a reliable source of information is at hand, clients would need to go back and forth from the real product to the place where its characteristics are presented, no matter whether they are written in a nearby sheet of paper or consulted in a smartphone. This process is less than optimal and can become tiresome after some repetitions. Besides, customers still need to interpret the meaning of the information and how it is linked to the product, a task that may happen to be too complicated for those that are not knowledgeable enough in the product space.

Our concept tackles these issues by using AR in a manner that consumers acquire relevant information just by looking at the products. The information is organized in attributes and categories as follows:

- **Attributes:** an attribute is formed by a name and a value. Attributes of each product are put together in categories and displayed anchored to a side of the physical object they belong to. The selection of an attribute shows a brief description that helps to understand its importance and, for those that are linked to some part of the product, the related

[1] engine.vuforia.com/

Figure 1: Overview of the prototype. The attributes of a product are presented on the left side, while recommended items appear on the right side.

Figure 2: The selection of an attribute highlights it and shows a brief description of its meaning and where it is located in the product.

physical part is highlighted to allow direct inspection (Fig. 2). Some attributes act as containers for others attributes (e.g. if a product is powered by a battery, the battery itself has attributes too); these "sub-attributes" are normally hidden, but they will appear if the attribute that contains them is selected (Fig. 3).

Categories: a category refers to a broader, significant aspect of the product (e.g. performance or required maintenance) and receives a score based on the values of the attributes that it encloses. The use of categories is specially significant for technical products, because they usually have many attributes that would overcrowd the display if no filtering

Figure 3: Attribute that contains sub-attributes. Bookmark (top) and Critique (right) buttons are shown as well.

means were provided. By accessing a category, the attributes of the product that have an effect on that specific aspect are revealed.

This design works towards making the information more accessible: the different categories help to create relationships between attributes, while the scoring system, attribute explanations and their linked physical parts give an insight of the product's qualities that is understandable even by clients who lack the required knowledge.

3.2 Product comparison

Comparing is regarded as a basic cognitive activity and holds great relevance in terms of understanding, exploring and evaluating our surroundings [14]. But comparing involves a mental effort and becomes harder the more information is required (e.g. when more than only two options are to be compared) as consequence of the limitations of short-term memory [2]. Once again, AR appears to tick all the right boxes: the mental workload of retaining information can be alleviated thanks to the utilization of spatial superimposition via AR, which eliminates short term memory demands[39].

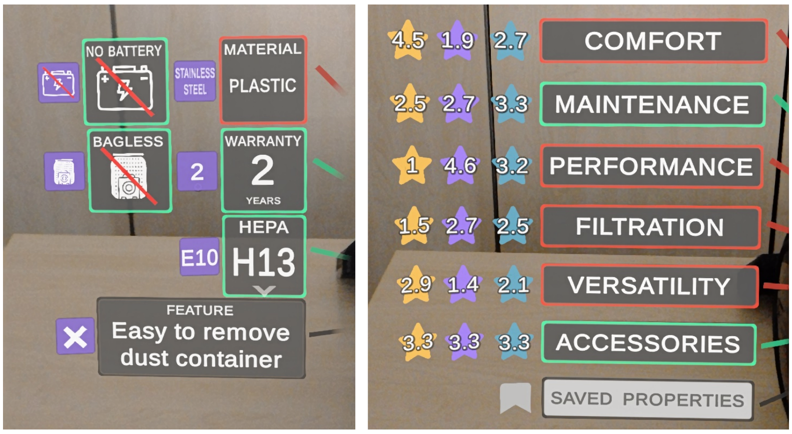


Figure 4: Attributes and categories during the comparison of 2 and 3 products respectively. Values belonging to different items are shown in the colour that has been assigned to them (yellow, purple and blue in the example; in this case, blue is the colour of the product the user is looking at when in the categories view). Selected products, and their assigned colours can be consulted on the right side of the UI (under the recommended items section), which is not shown here. Those attributes in the current product that are better than in the other compared items appear in green; otherwise, the are highlighted in red.

Considering how natural it is for consumers to compare different products before taking a decision, it only seems reasonable to at least explore the advantages that supporting the comparison of physical items may bring to the in-store context. Our approach aims to ease the comparison process and allow users to visualize differences between products regardless of the quantity of their attributes or their location within the shop.

Online stores already offer comparison tools that typically rely on the use of side-by-side tables of product attributes. This is a convenient method for online retailers where there is no physical item that the user can examine. However, as it has been mentioned in the previous section, using a similar solution in a physical store would add the extra effort of going back and forth from real products to the place where the comparison of their attributes is being displayed. Moreover, and in relation to our approach, showing a simple table of attributes on an AR display would not make much sense: in many cases the table would end occupying too much screen space, covering the real world (and products) behind it and giving an impression of disconnection between physical and digital elements. This leads to a waste of the potential of the technology and the rise of questions about why to use AR when more traditional means can achieve better results.

Keeping all the aforementioned points in mind, our approach makes use of the already explained categorization and positioning of product attributes for presenting the comparison. A product enters into comparison mode after being selected by using the "tap" gesture on it. The comparison takes place when two (or three) items are selected at the same time. The following elements are part of the comparison process:

Side by side values: attribute values of each one of the selected products are shown together, keeping the information attached to the product and organized in the same manner as when the comparison was not yet enabled. Values of different products are distinguished by highlighting them with a specific colour. Features not included in one of the products but appearing in the others are added to the former with a tag that indicates that they are "not included".

Visual aids to support the identification of differences: attributes are evaluated by using the same rules that are applied for scoring categories. Based on the results, it is possible to know whether the current product has the best value for a particular attribute among all the chosen ones (bordered in green) or not (bordered in red), but only when such distinction makes sense (Fig. 4). Besides, the comparison of product measurements is shown by superimposing them on a 1:1 3D scale representation that helps to better appreciate their relative dimensions.

Custom Category: an extra category called "Saved Properties" makes possible the combined comparison of multiple attributes that are scattered along different categories. It allows consumers to customize which attributes are shown within it and can be used to store only those that are relevant to their preferences.

Saved Products: up to seven products can be saved in digital form at any moment. They follow the client's movement and are placed slightly above the head in such a way that they can be selected and deselected even when the physical product that they represent is far away.

By combining these elements, it is possible to keep all the information required to make a comparison always within the client's reach, even if only one of the products to be compared is nearby. Customers can save the products that they like while exploring the shop, bring their attributes with them, and start the comparison as they please. It has to be noted that saving attributes and products are actions that are not exclusive of the comparison mode and can be performed at any moment. Also, the process here explained has made reference only to physical-to-physical product comparison, but physical-to-digital comparisons are also possible, as it is mentioned in the following section.

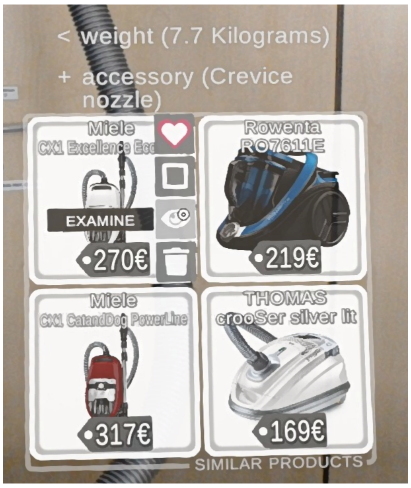


Figure 5: Recommended items. Critiqued attributes are listed above them.

3.3 In-store product recommendations

In our approach, recommendations are provided based on the physical product that is currently in the user's AR focus. Recommendations are calculated by using a content-based technique that takes the attributes of the product that has the attention of the client as an initial user preference model. Similar items are then retrieved from the database and four of them are displayed (Fig. 5). Two similarity scores are calculated: one based on functionality and another one on visual appearance. The functionality score only takes into account attributes that do not have an impact on the aspect of the product. On the other hand, the appearance score is obtained by mixing the similarity of visual-related attributes (e.g. colour, measures or material) and the outcome of comparing their product images against images of the base product. Image comparison is carried out by using DeepAI's API[2] which returns a value indicating how contextually similar they are (0 for identical images). When multiple images of a single product are available, only the lowest value is taken into account. This value is not calculated at runtime but stored in the database beforehand to avoid performance hiccups. The final set of recommendations is formed by the items with the highest scores, two of them based on functionality and two of them on appearance.

The initial set of recommendations can be further refined via critiquing (Fig. 5), which initiates a new recommendation process with a modified set of preferences that now includes the critiqued aspect. Categorical attributes are critiqued by telling the system whether it should be contained in the recommendations or not ("include this" or "exclude this") while for numerical ones it can be requested to consider higher or lower values. In any case, the critiqued properties do not act as a hard filter but as an added preference, thus recommendations are always retrieved. Critiqued properties are not universally shared among available physical products which means that they are set individually and create a unique preference model when joined to the base attributes of the product, thus obtaining distinctive recommendation sets.

[2] deepai.org/

Interaction-wise, other relevant aspects of the proposed concept are:

- Recommended items can be physical or digital, meaning that they may be accessible for inspection or not. This lets clients explore and choose items that are not in the shop, thus extending the catalogue and balancing the purchase options between online and physical retailing channels.
- Browsing the digital space is done by exploring the physical one. Recommended items change from product to product and are based on the specific item to which they are attached, thus users can find what they are looking for in the digital space by searching for similar products in the real world.
- Recommended items can be selected, saved, compared against physical ones or removed. The physical-to-digital comparison factor lets users experience their attributes by taking similar, real objects as a reference.
- Recommended items can be individually removed, which in turn brings forth new ones on their place.

These features provide a playground for clients to explore, learn and make decisions in a shopping situation. The whole buying process is supported: the system provides assistance from the information gathering phase to the point in which a final decision has to be made. Consumers with little knowledge about the product space can begin by exploring physical items in a natural manner; they learn about their attributes and other available purchase possibilities without any more hassle than looking at a product; product comparison is supported by the system, so that consumers do not need to remember attributes nor search for differences by themselves; clients can develop their own preference model that can be further elaborated by critiquing product attributes and obtain recommendations that adjust to it; finally, consumers are able to experience physical and digital products (to an extent) and take a more informed buying decision. The approach also provides a novel answer to open questions concerning the seamless integration of online and physical stores from a consumer's point of view.

4 EVALUATION

A study has been conducted to evaluate the validity of our approach and investigate the benefits of in-store recommending. The evaluation used only the system developed as no realistic baseline to compare against was available (a condition considering an online-only situation or a combined online-store scenario would have significant structural differences for it to be a comparable baseline).

4.1 Settings and experimental tasks

Three physical vacuum cleaner models were available (VC1, VC2, VC3), whose selection was done taking into account that they should cover different usage areas to let users explore a wide range of digital products through them, but remain similar enough to be compared. The database used to obtain the recommendations consisted of 100 vacuum cleaners.

During the study, a floating canvas shown via AR gave participants the information needed to complete the given tasks. Each participant had to solve two tasks concerning the search of an adequate product to match certain criteria. More specifically, each task asked to find products with the following characteristics:

Table 1: Short ResQue items

#	Question	mean	σ
1	The items recommended to me match my interests.	3.5	1.08
2	The recommender systems helped me discover new products.	4.1	0.99
3	The items recommended to me are diverse.	4.1	1.10
4	The layout and labels of the recommender interface are adequate.	4.0	1.05
5	The recommender explains why the products are recommended to me.	2.7	1.49
6	The information provided for the recommended items is sufficient for me to make a purchase decision.	3.9	0.87
7	I found it easy to tell the system what I like/dislike.	3.3	1.16
8	I became familiar with the recommender system very quickly.	4.1	0.87
9	I feel in control of modifying my taste profile.	3.7	1.49
10	I understood why the items were recommended to me.	3.5	1.17
11	The recommender helped me find the ideal item.	3.5	1.17
12	Overall, I am satisfied with the recommender.	3.9	1.19
13	The recommender can be trusted.	3.8	1.03
14	I will use this recommender again.	4.1	1.28
15	I would buy the items recommended, given the opportunity.	3.4	1.50

Task A High suction power and air flow values; moderate weight; attachments suitable for house and car cleaning; price under 200€.

Task B Small size and easy to store; good filtering system, appropriate for allergic persons; can handle pet hair; easy maintenance and handling.

It was taken into account that for each task at least one of the physical products could be a suitable choice.

4.2 Method

A total of 10 participants [3] (4 female, average age of 28.1, σ 4.06) took part in the experiment, 9 of whom had a strong technical background (3 Computer Science students, 5 PHD students and 1 telecommunications engineer in the industrial sector). Each participant was taught basic HoloLens usage and the main features of the prototype. After a brief time to let them get used to it and solve their questions, they were told to follow the instructions given by the application. Tasks were shown sequentially (but their order of appearance was counterbalanced between subjects), and after each of them a 3-item questionnaire was presented, treating aspects such as *purchase confidence* and *helpfulness of physical items.* After both tasks were completed they filled another questionnaire to assess the recommender systems' quality of user experience (*ResQue* [34]) and system-related items measuring the constructs *usefulness*, *decision-making* and *attractiveness.* In addition, task completion times and other empirical variables were measured.

4.3 Results

Items of the ResQue questionnaire are listed in Table 1. Most items show scores above 3.5, showing that users tended to rate positively the implemented recommender system in general, especially in those aspects concerning the novelty of the recommendations (items 2 and 3), perceived ease of use (8) and use intention (14). However, the system is lacking when it comes to the explanation of its results (5).

Table 2 shows the outcome of the system-related questions. They were also positively rated overall, but the highest scores were given to the preference of usage of the system over a traditional comparison tool (item 1) and participants' inclination to use the system if available (7). The two items concerning the helpfulness of physical products (2 and 3) were also rated favourably. However, participants seemed to encounter difficulties to find exactly what they wanted (item 4). Participants were also asked what they would choose if a salesperson and the system were both available. 2 of them would only use the system, 1 affirmed that he/she would prefer to only receive advice from the salesperson, and the remaining 7 would combine both asking for advice and using the system.

Items in Table 3 were answered by participants after each completed task (thus, twice per user of the system). They received high scores too, which suggests that the comparison function and the presence of physical products were perceived as helpful. Users also expressed confidence in their final choice.

Most participants said they felt able to use the system competently after having completed a first task fully. Despite completion times being shorter for the second task (Table 4), there were increments in the average number of critiqued, highlighted and bookmarked properties, as well as how often a category was changed and how many times a participant read the description of an attribute; user's attention moved from one product to another very consistently between tasks, and participants performed more interactions from a digital product to another digital product than from physical to physical or physical to digital ones.

Regarding the selection of a suitable product for each task, a digital one was chosen as the best fitting option in 18 occasions, while physical products were selected 2 times as the final choice (Table 5). For task A most final choices were either VC1 or an item recommended for VC1 or VC3, but never one of the recommendations based on VC2. Similarly, Task B was solved by choosing either VC2 or a recommended item based on VC1 or VC2, never for VC3.

[3] the number of participants needed to be limited because the study was conducted during the COVID-19 pandemic

Table 2: System-related items

#	Question	mean	σ
1	If I had the choice, I would prefer the proposed system instead of a traditional web-based product comparison tool.	4.10	1.97
2	I find that having physical products in front of me made it easier to make a decision.	3.90	1.10
3	I find it helpful/beneficial that I have the possibility to see/touch the products.	3.90	1.37
4	I found it easy to explore the product attributes and find what I was looking for.	2.80	1.03
5	I found it easy to compare the characteristics of different products.	3.70	1.05
6	I believe that I can make a faster buying decision by using the system than by using a more traditional mean (e.g. reading their attributes in a sheet of paper next to the physical products or consulting a salesperson).	3.50	1.17
7	If a store would offer this augmented reality application for a product I am interested in, I would use it.	4.4	0.69

Table 3: Within-subjects questionnaire

#	Question	mean	σ
1	I found it helpful to directly compare product features next to the physical product.	4.25	0.91
2	The physical product shown helped me to form an opinion about the products available online.	4.40	0.68
3	I am confident the product finally chosen would fulfil the requirements described in the task.	3.80	1.00

Table 4: Empirical data collected during the study

	first task		second task		overall	
	mean	σ	mean	σ	mean	σ
completion time (minutes)	14.8	6,20	8.28	1.8	11.54	5.56
frecuency of physical to physical product switches (switches/min)	2.46	2.48	2.59	1.79	2.52	2.10
frecuency of physical to digital product switches (switches/min)	6.85	3.69	8.21	4.91	7.53	4.27
frecuency of digital to digital product switches (switches/min)	11.57	4.81	11.33	5.99	11.45	5.27
products saved	3.00	1.32	2.00	1.00	2.50	1.24
properties critiqued per minute	0.56	0.34	0.72	0.45	0.64	0.39
properties highlighted per minute	1.99	1.01	2.65	1.48	2.32	1.25
properties bookmarked	2.00	1.14	2.22	3.07	2.11	2.32
property description readings per minute	0.60	0.40	0.75	1.5	0.67	1.06
category changes per minute	2.08	1.03	2.41	1.16	2.24	1.08

Table 5: Final choice by type (digital or physical) and item on which the recommendation was based. Task A/B refers to the description of the task, not their order.

	Based on VC1	Based on VC2	Based on VC3	Physical Product	Digital Product
Task A	7	0	3	1 (VC1)	9
Task B	3	7	0	1 (VC2)	9

Furthermore, 4 times out of 10 the same digital vacuum cleaner was selected among the 100 available as a solution for task A.

4.4 Discussion

The results suggest that participants considered the hybrid physical-online approach and the comparison and recommendation functions helpful. However, the system needed initial learning as can be seen in the more frequent use of some functions in the second task. Critiquing, for example, was used 28% more often in task 2.

Concerning the effectiveness of product recommendations via AR (RQ1), results of the ResQue items suggest a tendency towards a positive user perception of the implemented recommender system. Discovering new, diverse products has been relatively well rated, which may be the consequence of joining digital-product filtering through real-world exploration (selecting a physical product limits the digital space to only the similar ones) with critiquing techniques and the fact that recommendations where not only based on technical attributes but also on visual similarity (thus providing a more diverse set of recommendations). The intuitive process of discovering and filtering the digital space by exploring the real world could also be the reason behind participants generally finding easy to become familiar with the system, even with the added complexity that using a HMD may bring. Participants also seem to prefer the AR system over a traditional web-based one, which could be explained precisely by what makes both types of systems different, that is, the presence of real objects. This is supported by how highly regarded were the items of the questionnaire that deal with the helpfulness of having access to physical objects for both comparing and forming an opinion about products and their attributes as well as making a final purchase decision. Although the functions of the system were generally considered helpful, recommendation explanations and attribute exploration were scored lowest in the responses. Admittedly, recommendations have no more explanation than being linked to a physical product and being modified after a new critique is done, which is sufficient to understand the low scoring in that regard. However, the attribute exploration issue is not exclusive of the concept here presented nor recommender systems in general, and has more to do with the capabilities of AR technology when it comes to navigate through large attribute sets. Possibly, easier forms of presenting attributes and comparisons may become feasible when AR glasses will offer a larger viewing angle or if a different AR platform is used (e.g. smartphones, although they have their own challenges and limitations).

Regarding the possible implications of using AR product displays on the development of user preferences (RQ2), having access to physical products appears to be beneficial to some extent for understanding their properties and extrapolating them to the non-physically-present ones. Participants addressed this aspect directly in the survey, where the presence of physical products was judged mostly as helpful when making a decision in the digital sphere. Their perceptions in that regard are supported by how confident they were when assessing the suitability of the chosen products despite having selected digital ones for the most part.

Lastly, when it comes to possible anchoring or priming effects and the role that physical products play over the exploration of digital ones (RQ3), there appears to be a connection between what physical products are available and how users explore the digital space. The evaluation showed that users based their final choice on the physical item preferred for the task given and the recommendations that were provided for this physical item, mostly ignoring other online items. As has already been mentioned, selecting a physical product apparently acted as a filter for the online product space which suggest the existence of anchoring and priming effects.

4.4.1 Limitations. Participants of the study may not be representative of the population targeted by the concept due to their low number and strong technical background. For practical reasons there was only a limited number of physical products and not all product categories were properly represented , which would not have been the case in a real world scenario. This has an obvious impact in how often digital products were chosen over physical ones, which could have been very different if a larger variety of physical items were available. The usage of a HMD instead of a more conventional device may have had an impact on participants' perception of the system due to its novelty, thus possibly influencing the results. The interpretation of the results is also limited by the lack of a baseline against which to compare them and caution is required when interpreting an apparently positive finding. Lastly, the preferences implied by the task scenarios provided may have limited the users' need to engage more deeply in developing their own preferences. These aspects require a more careful exploration in further research.

5 CONCLUSIONS AND FURTHER RESEARCH

In this paper, an approach to in-store product recommendations provided via augmented reality is presented. It is built under the hypothesis that the creation of a mental preference model in a physical buying environment can be alleviated by having access to product information and recommendations, while at the same time the navigation of the digital space is improved by taking advantage of real world exploration. Furthermore, when the recommendations provide items that are not physically present in the store (taken from a digital catalogue), the inspection of other, similar items that are physically accessible could have an effect on how clients perceive the digital ones. In the implemented prototype for AR head mounted displays, clients can obtain product information of physical vacuum cleaners by just looking at them. The augmentations show product attributes and their meaning, as well as where the related parts are located in the product. Product comparison is supported by the inclusion of visual aids to directly perceive differences between them, which aims to mitigate the limitations of short-term memory. Recommendations of similar products (both real and digital) are provided for each physical one, whose outcome can be influenced by critiquing its attributes.

During the performed user study, the system was positively rated and perceived as useful and intuitive. The selection and exploration of products was influenced by the presence of physical items. The digital space was browsed based on the attributes of real objects, since participants focused on the recommendations given for specific vacuum cleaners (the ones that they considered more fitting for the given tasks). Physical products were predominantly regarded as helpful for forming an opinion of the ones available only in digital form. However, all these results should be taken with reservation due to the various limitations of the study in terms of the number of participants, the lab setting and the lack of a baseline. Altogether there is enough evidence to at least consider this to be a viable approach worth to be further explored. Future work will focus on consolidating the results presented here and on obtaining a deeper understanding of preference construction and the role of anchoring and priming effects in a hybrid physical-online setting. It is also in our scope to study new methods for combining appearance and function based recommendations and how different types of clients may react to them.

ACKNOWLEDGMENTS

This work was supported by European Regional Development Fund, project "FairWays" [EFRE-0801448].

REFERENCES

[1] Junho Ahn, James Williamson, Mike Gartrell, Richard Han, Qin Lv, and Shivakant Mishra. 2015. Supporting Healthy Grocery Shopping via Mobile Augmented Reality. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 1s (Oct. 2015), 16:1–16:24. https://doi.org/10.1145/2808207
[2] G.A. Alvarez and P. Cavanagh. 2004. The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science* 15, 2 (feb 2004), 106–111.
[3] Clemens Arth, Raphael Grasset, Lukas Gruber, Tobias Langlotz, Alessandro Mulloni, and Daniel Wagner. [n.d.]. The History of Mobile Augmented Reality. ([n. d.]). arXiv:1505.01319v3 [cs.HC]
[4] Marie Beck and Dominique Crié. 2018. I virtually try it … I want it ! Virtual Fitting Room: A tool to increase on-line and off-line exploratory behavior, patronage and purchase intentions. *Journal of Retailing and Consumer Services* 40 (jan 2018), 279–286. https://doi.org/10.1016/j.jretconser.2016.08.006

[5] Norbert Beck and David Rygl. 2015. Categorization of multiple channel retailing in Multi-, Cross-, and Omni-Channel Retailing for retailers and retailing. *Journal of Retailing and Consumer Services* 27 (nov 2015), 170–178. https://doi.org/10.1016/j.jretconser.2015.08.001

[6] Dirk Bollen, Bart P. Knijnenburg, Martijn C. Willemsen, and Mark Graus. 2010. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*. ACM Press. https://doi.org/10.1145/1864708.1864724

[7] Francesca Bonetti, Gary Warnaby, and Lee Quinn. 2018. Augmented Reality and Virtual Reality in Physical and Online Retailing: A Review, Synthesis and Research Agenda. In *Augmented Reality and Virtual Reality: Empowering Human, Place and Business*, Timothy Jung and M. Claudia tom Dieck (Eds.). Springer International Publishing, Cham, 119–132. https://doi.org/10.1007/978-3-319-64027-3_9

[8] Dries Buytaert. 2018. Shopping with augmented reality. https://dri.es/shopping-with-augmented-reality. [Online; accessed 2020-05-22].

[9] Federica Caboni and Johan Hagberg. 2019. Augmented reality in retailing: a review of features, applications and value. *International Journal of Retail & Distribution Management* 47, 11 (Nov. 2019), 1125–1140. https://doi.org/10.1108/IJRDM-12-2018-0263

[10] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui. 2017. Mobile Augmented Reality Survey: From Where We Are to Where We Go. *IEEE Access* PP, 99 (2017), 1–1.

[11] Edmanuel Cruz, Sergio Orts-Escolano, Francisco Gomez-Donoso, Carlos Rizo, Jose Carlos Rangel, Higinio Mora, and Miguel Cazorla. 2019. An augmented reality application for improving shopping experience in large retail stores. *Virtual Reality* 23, 3 (Sept. 2019), 281–291. https://doi.org/10.1007/s10055-018-0338-3

[12] Mussweiler Englich and T Mussweiler. 2016. Anchoring effect. *Cognitive Illusions: Intriguing Phenomena in Judgement, Thinking and Memory* 223 (2016).

[13] Fei Gao and Xuanming Su. 2019. New Functions of Physical Stores in the Age of Omnichannel Retailing. In *Operations in an Omnichannel World*, Santiago Gallino and Antonio Moreno (Eds.). Springer International Publishing, Cham, 35–50. https://doi.org/10.1007/978-3-030-20119-7_3

[14] Dedre Gentner and José Medina. 1997. Comparison and the Development of Cognition and Language. *Cognitive Studies* 4, 1 (March 1997), 1_112–1_149.

[15] Francisco Gutiérrez, Nyi Nyi Htun, Sven Charleer, Robin De Croon, and Katrien Verbert. 2019. Designing Augmented Reality Applications for Personal Health Decision-Making. https://doi.org/10.24251/HICSS.2019.212 Accepted: 2019-01-02T23:56:24Z.

[16] Gerald Häubl and Kyle B Murray. 2003. Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents. *Journal of Consumer Psychology* 13, 1-2 (2003), 75–91.

[17] Christian Homburg, Michael Müller, and Martin Klarmann. 2011. When Should the Customer Really be King? On the Optimum Level of Salesperson Customer Orientation in Sales Encounters. *Journal of Marketing* 75, 2 (mar 2011), 55–74. https://doi.org/10.1509/jm.75.2.55

[18] Toby Hopp and Harsha Gangadharbatla. 2016. Novelty Effects in Augmented Reality Advertising Environments: The Influence of Exposure Time and Self-Efficacy. *Journal of Current Issues & Research in Advertising* 37, 2 (July 2016), 113–130.

[19] Elodie Huré, Karine Picot-Coupey, and Claire-Lise Ackermann. 2017. Understanding omni-channel shopping value: A mixed-method study. *Journal of Retailing and Consumer Services* 39 (Nov. 2017), 314–330. https://doi.org/10.1016/j.jretconser.2017.08.011

[20] Stefanus Jasin, Amitabh Sinha, and Joline Uichanco. 2019. Omnichannel Operations: Challenges, Opportunities, and Models. In *Operations in an Omnichannel World*, Santiago Gallino and Antonio Moreno (Eds.). Springer International Publishing, Cham, 15–34. https://doi.org/10.1007/978-3-030-20119-7_2

[21] Ana Javornik, Yvonne Rogers, Ana Maria Moutinho, and Russell Freeman. 2016. Revealing the Shopper Experience of Using a "Magic Mirror" Augmented Reality Make-Up Application. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*. ACM Press. https://doi.org/10.1145/2901790.2901881

[22] Jiyeon Kim and Sandra Forsythe. 2008. Adoption of Virtual Try-on technology for online apparel shopping. *Journal of Interactive Marketing* 22, 2 (jan 2008), 45–59. https://doi.org/10.1002/dir.20113

[23] Cenk Kocas. 2002. Evolution of Prices in Electronic Markets Under Diffusion of Price-Comparison Shopping. *Journal of Management Information Systems* 19, 3 (Dec. 2002), 99–119. https://doi.org/10.1080/07421222.2002.11045740 Publisher: Routledge _eprint: https://doi.org/10.1080/07421222.2002.11045740.

[24] Sören Köcher, Dietmar Jannach, Michael Jugovac, and Hartmut H Holzmüller. 2016. Investigating Mere-Presence Effects of Recommendations on the Consumer Choice Process.. In *IntRS@ RecSys*. 2–5.

[25] Panos Kourouthanassis, Diomidis Spinellis, George Roussos, and George M Giaglis. 2002. Intelligent cokes and diapers: MyGrocer ubiquitous computing environment. In *First International Mobile Business Conference*. 150–172.

[26] Sören Köcher, Michael Jugovac, Dietmar Jannach, and Hartmut H. Holzmüller. 2019. New Hidden Persuaders: An Investigation of Attribute-Level Anchoring Effects of Product Recommendations. *Journal of Retailing* 95, 1 (mar 2019), 24–41. https://doi.org/10.1016/j.jretai.2018.10.004

[27] Georg Lackermair, Daniel Kailer, and Kenan Kanmaz. 2013. Importance of Online Product Reviews from a Consumer's Perspective. https://doi.org/10.13189/aeb.2013.010101

[28] Kelvin J. Lancaster. 1966. A New Approach to Consumer Theory. *Journal of Political Economy* 74, 2 (April 1966), 132–157.

[29] Tony Liao. 2014. Augmented or admented reality? The influence of marketing on augmented reality technologies. *Information, Communication & Society* 18, 3 (dec 2014), 310–326. https://doi.org/10.1080/1369118x.2014.989252

[30] Tariq Masood and Johannes Egger. 2019. Augmented reality in support of Industry 4.0—Implementation challenges and success factors. *Robotics and Computer-Integrated Manufacturing* 58 (aug 2019), 181–195. https://doi.org/10.1016/j.rcim.2019.02.003

[31] Eleonora Pantano. 2014. Innovation drivers in retail industry. *International Journal of Information Management* 34, 3 (jun 2014), 344–350. https://doi.org/10.1016/j.ijinfomgt.2014.03.002

[32] Young Park and Ulrike Gretzel. 2010. Influence of consumers' online decision-making style on comparison shopping proneness and perceived usefulness of comparison shopping tools. (2010).

[33] John W Payne, James R Bettman, and Eric J Johnson. 1992. Behavioral decision research: A constructive processing perspective. *Annual review of psychology* 43, 1 (1992), 87–131.

[34] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems - RecSys '11*. ACM Press. https://doi.org/10.1145/2043932.2043962

[35] Zulqarnain Rashid, Rafael Pous, Joan Melià-Seguí, and Marc Morenza-Cinos. 2014. Mobile Augmented Reality for Browsing Physical Spaces. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. ACM, New York, NY, USA, 155–158. https://doi.org/10.1145/2638728.2638796 event-place: Seattle, Washington.

[36] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender Systems: Introduction and Challenges. In *Recommender Systems Handbook*. Springer US, 1–34. https://doi.org/10.1007/978-1-4899-7637-6_1

[37] J. Ben Schafer, Joseph Konstan, and John Riedi. 1999. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce - EC '99*. ACM Press. https://doi.org/10.1145/336992.337035

[38] Anne R. Smink, Sanne Frowijn, Eva A. van Reijmersdal, Guda van Noort, and Peter C. Neijens. 2019. Try online before you buy: How does shopping with augmented reality affect brand responses and personal data disclosure. *Electronic Commerce Research and Applications* 35 (may 2019), 100854. https://doi.org/10.1016/j.elerap.2019.100854

[39] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the conference on Human factors in computing systems - CHI '03*. ACM Press.

[40] Miguel Torres-Ruiz, Felix Mata, Roberto Zagal, Giovanni Guzmán, Rolando Quintero, and Marco Moreno-Ibarra. 2020. A recommender system to generate museum itineraries applying augmented reality and social-sensor mining techniques. *Virtual Reality* 24, 1 (March 2020), 175–189. https://doi.org/10.1007/s10055-018-0366-z

[41] Endel Tulving and Daniel L Schacter. 1990. Priming and human memory systems. *Science* 247, 4940 (1990), 301–306.

[42] A. Tversky and D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (sep 1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

[43] Pasi Välkkynen, Alain Boyer, Timo Urhemaa, and Riku Nieminen. 2011. Mobile augmented reality for retail environments. In *Proceedings of Workshop on Mobile Interaction in Retail Environments in Conjunction with MobileHCI*.

[44] Peter C. Verhoef, P.K. Kannan, and J. Jeffrey Inman. 2015. From Multi-Channel Retailing to Omni-Channel Retailing. *Journal of Retailing* 91, 2 (jun 2015), 174–181. https://doi.org/10.1016/j.jretai.2015.02.005

[45] Felix von Reischach, Dominique Guinard, Florian Michahelles, and Elgar Fleisch. 2009. A mobile product recommendation system interacting with tagged products. In *2009 IEEE International Conference on Pervasive Computing and Communications*. IEEE. https://doi.org/10.1109/percom.2009.4912751

[46] Frank E. Walter, Stefano Battiston, Mahir Yildirim, and Frank Schweitzer. 2011. Moving recommender systems from on-line commerce to retail stores. *Information Systems and e-Business Management* 10, 3 (mar 2011), 367–393. https://doi.org/10.1007/s10257-011-0170-8

[47] Toñita Perea y Monsuwé, Benedict G.C. Dellaert, and Ko de Ruyter. 2004. What drives consumers to shop online? A literature review. *International Journal of Service Industry Management* 15, 1 (feb 2004), 102–121.

[48] Wei Zhu and Charles B. Owen. 2008. Design of the PromoPad: An Automated Augmented-Reality Shopping Assistant. *JOEUC* 20 (2008), 41–56.

[49] Jesús Omar Álvarez Márquez and Jürgen Ziegler. 2019. Augmented-Reality-Enhanced Product Comparison in Physical Retailing. In *Proceedings of Mensch und Computer 2019 (MuC'19)*. Association for Computing Machinery, Hamburg, Germany, 55–65. https://doi.org/10.1145/3340764.3340800

Paper VII. Acceptance of an AR-Based In-Store Shopping Advisor - the Impact of Psychological User Characteristics

Reproduced with permission from Springer Nature: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Acceptance of an AR-Based In-Store Shopping Advisor - the Impact of Psychological User Characteristics*. In *Human-Computer Interaction – INTERACT 2021*, Lecture Notes in Computer Science, pp. 457–479.

https://doi.org/10.1007/978-3-030-85623-6_28

Acceptance of an AR-Based In-Store Shopping Advisor - the Impact of Psychological User Characteristics

Jesús Omar Álvarez Márquez(✉) and Jürgen Ziegler

University of Duisburg -Essen, Forsthausweg 2, 47057 Duisburg, Germany
{jesus.alvarez-marquez,juergen.ziegler}@uni-due.de

Abstract. We present a study on the acceptance of augmented reality-based product comparison and recommending in a physical store context. An online study was performed, in which a working prototype for head-mounted displays, developed in previous research, was used to showcase the concept. The survey included questionnaires to assess shopping behaviour, decision styles and propensity to adopt new technologies of the participants. A cluster analysis of these psychological traits reveals the existence of different types of customers, who also differ on their assessment of the system. While the technology adoption propensity index is the better predictor of the acceptance of an augmented reality shopping advisor, the results suggest that factors such as the user's previous experience, a high experiential chronic shopping orientation, or an intuitive decision style have a significant impact on it as well. Thus, predicting user acceptance solely based on one of the investigated psychological traits may be unreliable, and studying them in conjunction can provide a more accurate estimation.

Keywords: Technology acceptance · Augmented reality · Retailing · Shopping advisors

1 Introduction

AR technology has made considerable advances in recent years [7], making it more readily available in a wider range of domains. Its usage has been successfully implemented in industry, specially concerning areas such as quality control, training and assistance in complex tasks [36]. AR is being well regarded in entertainment and marketing spheres too, due to the possibilities that it offers in terms of consumer engagement [14]. However, the use of AR in retailing is still scarce and most of the time e-commerce is the centre of attention, while physical retailing is left aside [34]. Bringing AR to physical stores requires finding more u-tilitarian uses for it [46], as well as suitable scenarios and proper visualization

Supported by European Regional Development Fund, project "FairWays" [EFRE-0801448].

Published by Springer Nature Switzerland AG 2021
C. Ardito et al. (Eds.): INTERACT 2021, LNCS 12932, pp. 457–479, 2021.
https://doi.org/10.1007/978-3-030-85623-6_28

and interaction methods. Furthermore, it is essential that an AR-based solution brings clear added value in contrast to more traditional options in order to be acceptable.

With all of this in mind, a promising use of AR is that of supporting in-store product comparison. Physical retailing lacks the ease with which online stores provide their customers with plenty of useful data and shopping tools. AR could be used precisely for closing this gap by letting customers explore product attributes via augmentations, while also allowing their comparison by assisting clients in the process of finding and understanding their differences. A potential benefit of the approach is that users are free to inspect attributes directly on the physical object they belong to, which may be of help for obtaining a better understanding of its qualities and, thus, make a more satisfactory purchase decision. In combination with recommending functions, AR can build a bridge between physical stores and online shopping to create multi-channel options [27,55], more so if the recommendations include products from the vendor's online catalogue. A system like this offers the rarely seen combination of digital and physical items of the same type, where the characteristics of all of them are accessible through a unified medium. This aspect may also have an impact on the decision making process of users by influencing how they explore the digital space and learn about physical and digital items.

However, it is unclear whether the use of AR technology in such context is acceptable to all users and which psychological characteristics may determine acceptance and attractiveness. Different general attitudes towards new technologies and user-specific shopping and decision-making behaviours may influence the acceptance of an AR system in a shop environment. Such factors may be particularly relevant when the system involves wearing an AR headset, which is conspicuous and may attract other customers' attention. Although research on the topic of AR acceptance already exists and even spans through different disciplines [18,37], it approaches the investigation mostly from a technical angle (e.g. users proficiency, availability of learning tools or current reliability of AR) or contextual elements, but overlooks the involvement of other psychological factors and the interactions that may occur between them. Thus, the following research questions are raised:

RQ1 How useful do users consider the possibility to explore and compare products across physical and online spaces?

RQ2 What is the impact of individual and combined personal characteristics on the acceptance of AR-based support functions?

This paper makes a further contribution to existing literature by presenting a exploratory study where, unlike previous research, users are defined by a set of psychological traits. These traits are based on how clients make decisions in a shopping scenario with a heavy technological component; that is, by assessing their technology adoption propensity, decision-making styles and shopping orientation. Participants are then grouped into types of clients to study their acceptance of an AR shopping advisor running on a head mounted display, in an attempt to find out which characteristics are more significant and uncover possible interactions that may exist between them.

2 Related Work

2.1 Use and Value of AR in Retailing

Studies show that AR has a positive impact on the shopping experience, particularly regarding customer satisfaction, consumer engagement and purchase intention [39,43]. Modern retailing can take advantage of AR at various consumer touch points, supported on the exponential growth of mobile technology [28] (e.g. IKEA's popular app [9]). Previous research shows that Mobile AR apps are perceived as valuable in retail contexts and provide benefits beyond the regular shopping experience [17]. Furthermore, due to its ability to merge digital and physical worlds together, AR acts as an enabler of omni-channel experiences by supporting the seamless integration of the different retailing channels [10,17,23]. Thus, it seems important for retailers to, at least, consider the adoption of AR-based experiences. However, the adoption of AR also presents its own challenges, such as taking the risk of its implementation, the initial investment in new technology or the need of training employees [17].

Previous approaches to AR in physical retailing include the PromoPad [59], an early application capable of providing context-aware information; Välkkynen et al. [54] explore the possibility of visualizing the content of a package before opening it; Rashid et al. [44] combined RFID with AR to browse product shelves; Cruz et al. [16] created an AR mobile application for retail stores that detects where the user is located and provides guidance to the item that the user is looking for. As of today and in terms of commercial success, virtual try-on [28] (or "magic-mirrors") are the most widely spread implementation of AR in physical contexts.

2.2 Shopping Advisors

Shopping advisors are very common in online settings, including features such as comparison tools [29,41], customer reviews and ratings [31] and product recommendations [50]. Per contra, it is difficult to find such elements in physical stores. An approach that brings such functionality into physical retailing can be found in Kourouthanassis et al. [30], where the authors present a system able to automatically create and keep track of a shopping list, and offer product information and personalized recommendations of promotions. APriori [47] is another example of a system that provides in-store product data, recommendations and user ratings.

Concerning AR, it has been stated that the technology offers improved search of information at the point of sale [52] and supports clients in making a purchase decision [15], characteristics that are desirable in a shopping advisor. Fully fledged AR shopping advisors are still rare, although some research exists on the topic: Ahn et al. [3] explore the benefits of using AR for product search in retail stores; Acquia Labs [13] developed a demo for a shopping assistant that provides, among other features, product information and in-store navigation support; Gutiérrez et al. [21] present a prototype for an AR shopping assistant that offers health-related information and discus different visualization layouts; Ludwig et al. [35] developed a working prototype to study the benefits of using AR to expose the underlying technical features of physical products. Commercial apps exist too: Aisle411 and Tango partnered to develop an app for Walgreens stores [1] that delivers product information, promotions and in-store navigation; or the Olai Skin Advisor [2], which offers recommendations of products after detecting the consumer's face skin conditions. Nonetheless, despite the existence of previous research on AR-based in-store shopping assistants, the combination of digital and physical products that can be seamlessly compared and recommended remains unexplored.

2.3 Acceptance of Augmented Reality Technology

Technology acceptance is defined by Dillon [19] as *"the demonstrable willingness within a user group to employ IT for the tasks it is designed to support"*. Despite what intuition might tell us based on that definition, the results provided by Roy et al. [49] suggest that technology readiness (i.e. an individual's propensity to embrace and use new technologies) may only influence customer acceptance towards smart retail technologies to some extent, that is, under certain conditions and for certain customers, while other factors such as perceived usefulness (PU), perceived ease of use (PEOU), and perceived adaptiveness play a larger role. Precisely, the review of existing literature [42] shows that one of the most widely used approaches to assess user acceptance of augmented reality in retail is the Technology Acceptance Model (TAM) [33], which considers that PU and PEOU are the main drivers of technology acceptance. The model has undergone several revisions through the years by both original and independent researchers [11], and it is often criticized because of its simplicity, which neglects the differences in decision-making and decision makers across technologies [8]. However, it is still a widely used model and considered valid for AR applications [18]. The extended versions of the TAM often discuss the addition of new factors such as perceived enjoyment (although the findings about its impact on user acceptance are conflicting) and perceived informativeness [24,42]. As a contrasting note, several authors opt for using flow theory instead [57], which focuses on the four dimensions of immersion, curiosity, fun and control.

Security and privacy aspects have been flagged as other relevant factors that influence the acceptance of AR technology, where previous literature [46,56] show that AR systems do not currently offer enough protection in that regard or, at the very least, do not sufficiently transmit the feeling of it.

When exploring the different factors involved in the acceptance of AR, existing studies mostly focus on the impact of aspects such as the characteristics of the technology (real or as perceived by users), psychological factors and environmental influences [18,42]. However, the existence of different types of consumers (defined by the combination of several of these elements), and how they may differ in their perception of AR in retailing settings, are questions that have been generally overlooked.

3 Research Questions

3.1 RQ1: How Useful Do Users Consider the Possibility to Explore and Compare Products Across Physical and Online Spaces?

Supporting the comparison process is a key component of the prototype that is evaluated here. This is justified by the great relevance of comparing in how human beings learn about their environment [20] and, consequently, the significant role that it plays in consumer behaviour, where comparing products is the most natural way to reach a purchase decision [32]. A previous evaluation of the system [6] suggested that combining digital recommendations and physical items may be beneficial for understanding the qualities of the not physically present products, due to the possibility of learning about them through the examination of the real ones. Moreover, it also seems to exist some connection regarding how users navigate the digital space and what products are physically available, in a way that digital items are intuitively filtered out by exploring only the recommendations provided for already suitable, physical products. These points indicate some potential benefits of offering such in-store services, but there is still a need to confirm these results by surveying a larger population sample.

3.2 RQ2: What Is the Impact of Individual and Combined Personal Characteristics on the Acceptance of AR-based Support Functions?

There is a research gap concerning how different decision-making-related psychological traits participate in a user's acceptance of in-store AR applications. To determine what these traits could be, we take the work by Alavi and Joachimsthaler [4] as reference, where the psychological variables involved in the acceptance of a decision support system are examined. *Cognitive style*, *personality traits*, *user situational variables* and *demographics* are identified as the most relevant factors. In the following, the measurement of each factor (as defined by Alavi and Joachimsthaler) is discussed.

Cognitive style refers to how information is processed and used. Different scales exist that allow its analysis, such as the Decision Styles Scale [22], which only requires 10 items to provide an outcome on two different scales (*rational* and *intuitive*); or the more complete and commonly used approach by Scott and Bruce [51], which distinguishes between *rational*, *avoidant*, *intuitive* and *dependent* decision-making styles, but at the expense of a greater number of items.

Personality traits are such as need for achievement, degree of defensiveness, locus of control or risk-taking propensity. Given that the matter at hand consists in the inclusion of a very intrusive technological component (a head mounted display) in a physical retailing context, we aim at the assessment of those traits involved in both technology adoption and shopping behaviour.

Concerning technology adoption, it has already been stated that the Technology Acceptance Model [33] is the most popular theory. The Technology Readiness

Index [40] is another well-known tool for measuring an individual's propensity to adopt and embrace new technology, and it focuses on four different dimensions that act as motivators (*optimism*, *innovation*) or inhibitors (*discomfort*, *insecurity*). Consistent with this idea, the Technology Adoption Propensity index [45] also considers the existence of positive (*optimism*, *proficiency*) and negative (*dependence*, *vulnerability*) attitudes in the assessment of technology acceptance, but uses a more contained set of items.

Among the alternatives for measuring consumer-related traits, the most prominently used is the Consumer Styles Inventory [53], which profiles individuals by analysing eight basic characteristics. Westbrook and Black [58] propose another widely used approach based on *hedonic* and *utilitarian* shopping motivations. A more recent study by Büttner et al. [12] discusses the creation of a 7-items long Chronic Shopping Orientation Scale, which aims at the prediction of the consumer's stable shopping disposition (*experiential* or *task-focused*).

Demographic data on gender, age and education can be easily collected. As for user-situational variables, the work by Alavi and Joachimsthaler [4] refers to user training and experience, which in our case could be associated with the user's previous knowledge about augmented reality.

Gathering information on each one of these personal characteristics would allow the investigation of their role in the judgement of a system like the one described in the following section. Furthermore, by uncovering possible relationships between these characteristics, it would be possible to determine the existence of distinguishable customer types and any variations in their acceptance of the system.

4 Description of the Prototype

A prototype for an AR-shopping advisor designed for Microsoft HoloLens was developed [5,6]. As of today, using a smartphone could have been a more practical approach. However, this research chose to use a head mounted display as AR enabler because of its growing relevance and availability, and its facility to provide a more engaging experience (which is specially relevant in shopping contexts) and allow for more interesting interaction possibilities (e.g. hands-free direct inspection of products).

Vacuum cleaners were chosen as the product domain, since they are common, technical commodity products. The approach, however, could have been applied to a wide range of products, so long as they are rich in attributes and their physical qualities are relevant for consumers. It is also necessary to keep in mind that this approach may not be best suited for standard shopping environments, but in stores whose activities include working as show-rooms: spaces where a carefully selected range of products are presented (usually specialized in a specific type of items), and where clients have enough space to wander around and freedom to examine them.

Fig. 1. Main view of the system. Attribute categories and recommendations are placed on the left and right sides of the product.

4.1 Access to Product Information

When a user looks to a product that is physically present in the store, relevant information is displayed surrounding it (Fig. 1). Clients can then move from one product to another to inspect them individually. Product information is organized in categories that group attributes based on their impact over a certain aspect. Within a category, the system shows the values of the attributes that belong to it. They can be selected, which shows extra information (Fig. 2).

Against the argument that similar results could be achieved by more standard means (e.g. a smartphone app with object recognition capabilities), an AR approach enables the inclusion of relevant spatial information for each attribute (i.e. where they are located), bringing into play a new layer of interaction between digital and physical worlds where their connexion is made more apparent. Such union should call for direct exploration and testing of physical items, while improving the understanding of their digitally displayed properties.

4.2 Product Comparison

Taking online comparison tools as a reference, the prototype lets users select up to three different products to be compared. When the user looks to one of the selected items, the attributes of the other one(s) (not directly in the client's line of sight) are shown next to the attributes of the former, in a side-by-side manner (but maintaining the same attribute organization and exploration methods that have been previously explained). Each product is assigned a specific colour that helps to distinguish their properties. The system highlights in green or red the

Fig. 2. Selecting an attribute shows its location on the product, a brief description and critiquing buttons (+/−).

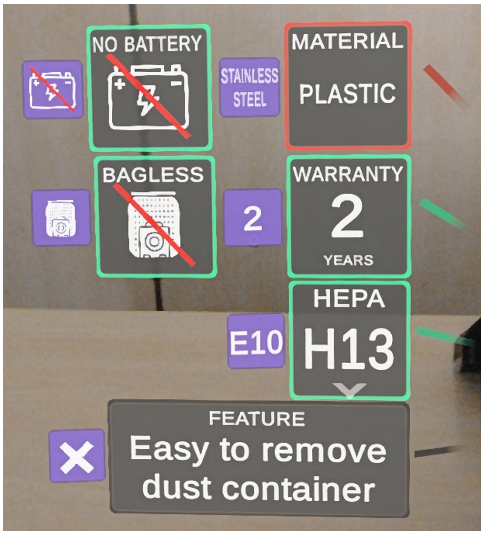


Fig. 3. Comparison view. Colour-coded attribute values (e.g. purple) belong to products not in the client's line of sight. (Color figure online)

attributes of the current product to easily identify in which ones it performs better or worse than the other selected items (Fig. 3).

4.3 Product Recommendation

Product recommendations are obtained likewise attribute information is retrieved: just by looking at a certain product. Items similar to the product directly in the line of sight of the client are shown next to it (Fig. 1). These recommendations can be directly compared without requiring users to find the real objects they represent. This allows the inclusion of recommendation of products that are not physically present at the store, effectively expanding the store's catalogue.

Recommendations can be modified, either by directly removing those that are not wanted (which brings new ones in their place) or by critiquing the attributes of the item for which they have been recommended (users can choose whether they seek for products with higher or lower values in a particular property).

5 Online Study

An online study has been conducted to assess the acceptance of an AR-based in-store shopping advisor designed for head-mounted displays and the specific functions described in the previous section. An online study design was chosen to get a broader feedback following an initial lab study[1] [6]. The goals of the study

[1] Conducting another, larger lab experiment was also considered, but that option had to be discarded due to restrictions related to the COVID-19 pandemic.

include defining different consumer profiles based on their propensity to adopt technology, shopping orientation and decision styles, and to analyse whether differences exist in their judgement of the system. Finding possible relationships between the psychological traits that characterize those profiles and their impact on the acceptance of the concept is also within the scope of this investigation.

5.1 Settings of the Study

Participants were asked to complete an online questionnaire with three distinct parts. The first one focused on determining the psychological profile of the participants and included questionnaires to that effect. Then, since participants could not experience the system by themselves, a video was included that had an initial part where the concept was introduced, followed by a thorough description of the prototype's functionality, showcased with real images of the system as seen from the user's point of view. The last part of the survey was comprised by a set of questions in relation to the aforementioned video with the purpose of collecting information about the acceptance of the prototype.

5.2 Method

A total of 63 participants (40 females, average age of 34.1, σ 12.29) took part in the study. They were recruited through the online platform *Prolific*[2] and received a monetary award of £1.50 after successfully completing a survey (as per the site's policy). As it can be seen in Table 1, the majority were residents of the United Kingdom, while most of them had achieved a master's degree level of education and worked either as employees or were self-employed. Furthermore, many of them reported to have limited knowledge about augmented reality and its possible applications (Table 2).

Table 1. Demographic information of the sample

Country	#
United Kingdom	43
United States	8
Ireland	4
Netherlands	2
Canada	2
Other	4

Education level	#
Secondary school	2
High school	9
Apprenticeship	1
Bachelors degree	19
Masters degree	32

Working status	#
Pupil/in school	3
Training/apprenticeship	1
University student	8
Employee	28
Civil servant	4
Self-employed	10
Unemployed	9

[2] www.prolific.co.

Table 2. Knowledge about augmented reality technology

I know nothing about it	8
I know the name, but not much more	21
I know a bit about it and its possible applications	30
I have followed its development and know it well	4
I know a lot and could be considered an expert in the field	0

The survey itself was hosted by *SosciSurvey*[3]. The first part of the survey comprised questionnaires that were to be used in the creation of the participant's psychological profile. Since different traits had to be measured, and to prevent response fatigue and collect more truthful answers [38], the length of the questionnaires included in the survey was an important factor when selecting them. Even though the chosen questionnaires might not offer as much information as others available, their combination should suffice to create a reliable user profile:

– *Technology Adoption Propensity* (TAP) index [45], a 14-items-long questionnaire that provides a score that represents how likely a person is to adopt new technology by considering four sub-scales: *optimism*, *proficiency*, *dependence* and *vulnerability.* A TAP score is equal to the sum of the average scores on each of the four factors, with inhibiting factors reverse coded. Each individual item is rated using a 5-point Likert scale.
– *Chronic Shopping Orientation* (CSO) scale [12], which assesses whether a person has a stable consumer disposition to shop under an experiential or a task-focused shopping orientation. It uses a 7-points Likert scale, ranging from *task-oriented* (lower values) to *experience-based* (higher values) shopping orientations.
– *Rational and Intuitive Decision Styles* (RDS and IDS) scale [22], which reflects the prevailing manner by which individuals make decisions. Each decision style is measured independently, hence two scores (in a 5-point Likert scale) are provided.

Following these questionnaires, a video that explains the concept and showcases the implemented prototype for Microsoft HoloLens was presented to participants[4]. After watching it, a final set of system-related questions were presented to assess the acceptance of the concept. These questions were designed to measure the constructs *perceived usefulness* and *perceive ease of use* (the two main constructs in TAM), extended with *decision-making support*, *hedonic motivation* and *intention to use*. Items regarding *social acceptance* and *privacy* were also part of it, which have been highlighted as inhibitors in the adoption of AR headsets [46].

[3] www.soscisurvey.de.
[4] The video is available on www.youtube.com/watch?v=k4nyTDQ-n7U.

In the following, to examine the overall acceptance of the concept and the usefulness of the AR functions provided (RQ1), the results obtained for the sample are examined as a whole. Afterwards, psychological data is used to find how different psychological traits influence the adoption of an in-store AR advisor, either individually or combined to define consumer types (RQ2).

5.3 General Results

Table 3 reports descriptive data obtained for each one of the investigated constructs and the items within them, which are relevant for determining the overall acceptance of the system and the usefulness of the combination of digital and physical elements (RQ1). *Perceived usefulness* and *decision support* received the highest scores among all the constructs, while *privacy* and *social acceptance* obtained the lowest ones. However, it has to be noted that all constructs fall into the positive side of the scale (their scores are higher than 3). Furthermore, although the construct *importance of physical items* appears in fourth position, the score for the single item *"Inspecting products will help me to make a more informed buying decision"* is among the most highly rated.

The data in Table 4 shows that providing *comparison support* is perceived as the most important feature of the system. *Product recommendations* and *access to a digital catalogue* are close to each other, while *interaction with physical items* falls a bit behind, in the last position. These results are in line with the scores given to the constructs, where *decision support* (comparing and recommending) obtained higher ratings than *importance of physical products*.

Finally, when having to choose between using the system or being assisted by sales personnel, 21 participants stated that they would use the system only; other 29 said that they would use the system first, and ask for support if required; 10 would first ask for support, and likely use the system afterwards; and the remaining 3 would not use the system at all.

An analysis of the Pearson's product-moment correlation between psychological traits and the questionnaire's constructs was performed to study the impact of individual personal traits on the acceptance of the system (RQ2). After the adjustment of the p-values by using the Benjamini-Hochberg procedure, only a moderately negative correlation between intuitive decision style and *social acceptance* ($r(89) = -.37, p < .05$) was found. Although the results hint at other possible correlations, they are not strong enough to be reported here.

5.4 Cluster Analysis

In order to find combinations of user characteristics that may have an effect on system acceptance (RQ2), it is first necessary to classify users into customer types, which allows the comparative analysis of their responses. In our case, this classification considers the scores that subjects obtained in TAP, CSO, RDS and IDS scales, that is: how likely they are to adopt a new technology, the way they approach shopping and how they make decisions. To this end, a two-step process was performed: first, data was classified through a hierarchical cluster analysis

Table 3. Overall results for the system-related items

Constructs and items from higher to lower mean	Mean	σ	95% CI Lower	95% CI Upper
Perceived Usefulness	**4.18**	**0.84**	**3.97**	**4.39**
I find the system will be useful	4.21	0.85	3.99	4.42
I believe that the use of Augmented Reality is beneficial in the given scenario (physical retailing)	4.16	0.90	3.93	4.39
Decision Support	**3.94**	**0.81**	**3.74**	**4.15**
I can have a better view of all the available choices with the help of the system	4.16	0.94	3.92	4.39
The system will help me to discover new products	4.13	0.92	3.89	4.36
By using the system, I think it will be easier to find an item that I like	4.00	0.90	3.77	4.23
This recommender system will increase my confidence in my selection/decision	3.98	1.04	3.72	4.25
By using the system, I think it will be easier to find an item to buy	3.86	1.00	3.61	4.11
Using the system will help me to make a decision more quickly	3.52	1.06	3.26	3.79
Interface Adequacy	**3.94**	**0.95**	**3.70**	**4.18**
The information provided for the recommended items will be sufficient for me to make a purchase decision	3.94	0.95	3.70	4.18
Importance of Physical Products	**3.78**	**0.70**	**3.60**	**3.95**
Inspecting products will help me to make a more informed buying decision	4.11	0.85	3.90	4.32
Interacting with physical products will help me to understand the features of similar, digital ones	3.89	0.81	3.69	4.09
I would use the system even if a digital catalogue were not available (the system would only recommend physically available products)	3.69	1.06	3.42	3.95
If the same catalogue is available at an online store and a physical one, I generally prefer to do the extra effort of travelling to the physical one to inspect the products by myself	3.43	1.10	3.15	3.71
Intention to Use	**3.67**	**0.95**	**3.43**	**3.91**
Assuming I had access to the system, I would likely use it	3.89	0.99	3.64	4.14
Being able to use the system will be a reason for choosing one store over another	3.57	1.20	3.27	3.87
When in a physical store, I would rather use this system than a more traditional web-based recommender	3.54	1.11	3.26	3.82
Perceived Ease of Use	**3.64**	**0.92**	**3.40**	**3.87**
Learning how to use the system will be easy for me	3.94	1.08	3.67	4.21
I think that the interaction with the system will be clear and understandable	3.70	1.01	3.44	3.95
I find the system will be easy to use	3.52	1.03	3.26	3.78
Interacting with the system will be an effortless task	3.38	1.07	3.11	3.65

(continued)

Table 3. *(continued)*

Constructs and items from higher to lower mean	Mean	σ	95% CI Lower	95% CI Upper
Hedonic Motivation	**3.61**	**1.16**	**3.32**	**3.90**
Using the system will be fun	3.71	1.18	3.42	4.01
Using the system will be entertaining	3.51	1.19	3.21	3.81
Privacy (reversed)	**3.25**	**1.14**	**2.97**	**3.54**
I would be concerned that my data is stored and used for other purposes	3.05	1.25	2.73	3.36
I would have privacy concerns if someone uses the glasses around me (e.g. when that person looks at me)	2.44	1.20	2.14	2.75
Social Acceptance (reversed)	**3.22**	**0.88**	**3.00**	**3.44**
Being able to share the experience with my shopping partner (e.g., we both see and interact with the same AR elements in real time) is relevant for deciding whether to use the system or not	3.35	1.17	3.06	3.64
I would not feel comfortable using the system while other people are around	2.64	1.24	2.32	2.95
I would find it annoying/irritating when other person uses the system nearby	2.35	1.21	2.05	2.65

based on average linkage between groups, which provided information on outliers and an initial distribution of participants; second, a K-means clustering analysis was conducted to confirm the results obtained in the first step. Since the variables involved have different scales, their z-scores were used for clustering purposes.

The outcome shows that four well-distinguishable groups can be identified, although 9 out of the total of 63 participants are considered as outliers and can thus not be classified. The silhouette scores [48] for each cluster are reported in Fig. 4, and the average results that each group obtained in the psychological tests are shown in Table 5. It is possible to identify what traits characterize each cluster by considering the relation between their means and the average of the sample (in the following, "group" and "cluster" are used interchangeably):

Table 4. Importance assigned by participants to each feature of the system (1–5 scale).

Feature	Mean	σ	95% CI Lower	95% CI Upper
Product comparison support	4.60	0.61	4.45	4.76
Access to product recommendations	4.33	0.78	4.14	4.53
Access to a digital catalogue	4.27	0.77	4.08	4.46
Interaction with physical products	3.94	1.03	3.68	4.20

Group 1 is composed of average technology users (relative to this sample). This group is the most experiential one when it comes to shopping and has an over the average intuitive decision-making style. It is also the largest group, gathers the youngest participants and is mostly composed of females.

Group 2 includes people with higher probabilities of adopting new technology, who also see shopping as a task-oriented experience. The group presents a notable polarization between rational and intuitive decision scales: their RDS is the highest among groups, while their IDS shows the lowest value.

Group 3 has the lowest probabilities of adopting new technology: its members do not perceive it as useful, are not proficient at it and believe themselves to be very dependant and vulnerable. Consequently, they know the least about AR. They also show the lowest rational decision style score.

Group 4 has a TAP value similar to that of group 2, but its CSO, RDS and IDS scores are more moderate, and has more knowledge of AR technology than any other cluster. The average age is higher than that of the other groups and it is the only one composed predominantly of male participants.

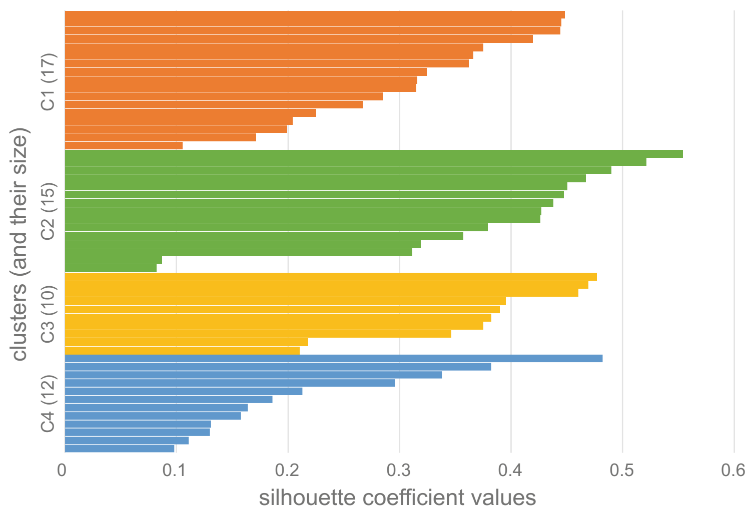


Fig. 4. Silhouette scores by clusters. Average silhouette width = 0.32.

Table 6 shows the scores obtained by cluster in each construct, which already suggests a difference in their assessment of the system. Although the ranking of the constructs is mostly maintained between groups (with perceived usefulness and decision support in the first places and privacy and social acceptance in the last), the range of scores differs: group 2 has the most positive view of the system and group 3 the most negative; group 1, on the other hand, is the most polarized (it provides some of the highest and some of the lowest scores of the sample), while, in contrast, group 4 shows a more uniform and moderate rating distribution.

Table 5. Psychological traits per group. From left to right: group size; percentage of females; average age; AR knowledge (on a 1–5 scale); TAP sub-scales (1–5): Optimism, Proficiency, Dependence, Vulnerability; total TAP score; Chronic Shopping Orientation (1–7); Rational and Intuitive Decision Styles (1–5); and the difference between them. Values of the psychological traits are relative to the average of the sample without outliers (last row of the table). Coloured cells identify those values where a group's mean noticeably differs from the total.

	#	Fem.	Age	AR Kno.	TAP Sub-scales Opt.	Pro.	Dep.	Vul.	TAP Score	CSO Score	RDS Score	IDS Score	RDS - IDS
G1	17	88%	28.53	-0.09	0.10	0.01	0.30	0.20	-0.39	1.23	0.05	0.58	0.72
G2	15	47%	34.53	0.09	0.06	0.35	-0.36	-0.52	1.28	-0.84	0.46	-0.91	2.63
G3	10	70%	33.70	-0.44	-0.30	-0.96	0.78	0.76	-2.80	-0.37	-0.50	0.22	0.54
G4	12	33%	37.75	0.39	0.03	0.36	-0.62	-0.27	1.29	-0.37	-0.24	0.13	0.88
			Avg.	2.44	4.27	3.68	2.95	3.60	13.40	3.58	4.25	3.00	1.25

Table 6. Differences in the assessment of the system among user groups.

	Group 1 Mean	σ	Group 2 Mean	σ	Group 3 Mean	σ	Group 4 Mean	σ	Total Mean	σ
Perceived Usefulness*	4.44	0.58	4.43	0.65	3.75	0.79	3.83	1.03	4.18	0.80
Decision Support	4.20	0.62	4.19	0.47	3.52	0.81	3.89	0.72	4.00	0.68
Interface Adequacy	4.06	0.90	4.27	0.88	3.40	0.84	3.92	0.79	3.96	0.89
Imp. of Physical Products	4.04	0.54	3.93	0.70	3.78	0.77	3.40	0.58	3.82	0.67
Hedonic Motivation*	4.06	1.00	4.00	1.02	3.00	1.00	3.54	0.81	3.73	1.02
Intention to Use*	3.98	0.69	4.11	0.87	3.27	0.94	3.28	0.84	3.73	0.89
Perceived Ease of Use	3.72	0.96	3.88	0.60	3.23	1.10	3.71	0.80	3.67	0.87
Privacy	3.09	1.08	3.77	1.05	2.90	1.02	3.38	0.12	3.31	1.06
Social Acceptance*	2.86	0.53	3.87	0.75	2.93	0.93	3.19	0.93	3.23	0.86

* *There is statistically significant difference between groups ($p < .05$)*

The analysis of the means by performing a Kruskal-Wallis test revealed statistically significant differences between clusters for *perceived usefulness*, *hedonic motivation*, *intention to use* and *social acceptance*. Subsequent Dunn-Bonferroni pairwise comparison tests reported a significant difference between groups 1–3 in terms of *hedonic motivation*, and groups 1–2, and 2–3 regarding *social acceptance* ($p < .05$ in all cases). In that regard, group 1 shows the highest *hedonic motivation* but, at the same time, it seems to be the most worried about *social acceptance*, in contrast to groups 3 (the least attracted by the system) and 2 (the least concerned about *social acceptance*). These results agree with the previously established negative correlation between intuitive decision style and *social acceptance* because, as a matter of fact, group 1 comprises the most intuitive participants. As for the remaining statistically significant constructs, post-hoc tests were not able to specify the groups for which the differences were significant, which suggests the existence of more complex relationships. Further testing of different group combinations showed that there is a statistically significant difference between the union of groups 1 and 2 and the union of groups 3 and 4 for *perceived usefulness* and *intention to use* ($p < .01$ in a Mann-Whitney U test).

Performing the same procedure over individual items of the questionnaire shows more significant differences, but most of them are in line with the test involving constructs and, therefore, are not reported here. However, some new disparities were found in relation to *decision support*, more specifically for *"This RS will increase my confidence in my selection/decision"* (between groups 1–3, $p < .05$) and *"I can have a better view of all the available choices"* (between groups 2–3, and 2–4, both $p < .05$).

Regarding what features are more important (Table 7), groups with higher TAP (2 and 4) seem to prioritize exploration-related functions (comparison and digital catalogue). On the other hand, group 3, the least technologically proficient and rational, finds product recommendations to be the most valuable feature. It also seems that the lower the TAP value is, the more importance is given to physical interaction with products.

Lastly, Table 8 shows the distribution of answers concerning users intention to use the system when sales personnel is also available. The responses are consistent with each group's assessment of the system. Interestingly, it is in group 4 (the one with the highest knowledge about AR) where, given the chance to use the system, one of its members would choose not to do so.

Table 7. Importance assigned by groups to each feature of the system (1–5 scale, 'R' is for rank).

Feature	Group 1			Group 2			Group 3			Group 4		
	m	σ	R	m	σ	R	m	σ	R	m	σ	R
Product comparison	4.59	0.51	1	4.73	0.46	1	4.40	0.70	2	4.67	0.65	1
Product recommendations	4.53	0.51	2	4.07	0.88	3	4.60	0.96	1	4.25	0.75	3
Digital catalogue	4.18	0.64	3	4.20	0.77	2	4.20	0.92	3	4.33	0.78	2
Int. with physical products	4.12	0.86	4	4.00	1.25	4	4.20	0.79	3	3.42	0.90	4

Table 8. Intention to use the system by group.

If i need support and the system is available ...	G1	G2	G3	G4
... I would use it to find what i need	6	7	1	4
... I would use the system and ask for assistance only in case of doubt	9	8	4	5
... I would first ask for assistance and likely use the system afterwards	2	0	5	2
... I would prefer to receive assistance from sales personnel only	0	0	0	1

5.5 Discussion

In relation to the usefulness of the hybrid physical-digital approach (RQ1), the feedback received on the AR-based shopping advisor can generally be considered positive. While it was not possible to assess the prototype against a baseline due to the lack of a system with comparable functionality, the scores received for the different functions and interaction methods are all in the positive range. The constructs *perceived usefulness* (considered one of the main drivers of user acceptance in conjunction with *perceived ease of use*) and *decision support* received particularly high ratings. Although interacting with physical products is overshadowed by the other features, it is still regarded as helpful to make a more informed buying decision.

As for the impact of the traits investigated here (RQ2), none of them appear to fully determine the acceptance of in-store AR-based advisors. In some cases, participants with apparently opposing psychological profiles seem to coincide in their assessment of the system. For instance, both persons with a heavily task-oriented CSO and a rational DS and those with an experiential CSO and an intuitive DS are more likely to perceive the system as useful and to use it. In contrast, those with a low TAP score who also have little knowledge about AR and those with a high TAP score and knowledgeable in AR technology are less inclined to its usage. However, their reasons for accepting or rejecting the system may as well be very different: perhaps they see a practical value in it, or think that it may offer an enjoyable experience; maybe they simply do not like technology in general or, if they do, they have had a higher exposure to AR already, in a way that they are less impressed by its novelty and more aware of its current limitations. In general terms, a lower TAP score appears to be the better predictor of the rejection of the system, but average to high TAP scores seem to be moderated by previous experience with AR technology. This may be a reflection of how quickly the novelty of AR fades away [25] and the difficulty of finding reasons to use it on a regular basis instead of a more traditional method [26]. Finally, a prominent experiential CSO and an intuitive DS seem to lead to a good acceptance of the concept too, although overexposure to AR could have the same moderating effect as with high TAP values.

The outcome of the study also indicates that rejection of the system due to privacy or social acceptance concerns is more likely to happen on subjects with higher intuitive decision styles and lower TAP scores. In spite of that, it must be noted that these worries were present in all groups to a certain degree. This finding agrees with previous research on the topic where privacy and social acceptance are identified as challenges in the adoption of AR [46, 56]. Some participants left comments where they defined the glasses as *"too creepy"* or would express their fear to *"be pick-pocketed or robbed when not on guard"*. These worries are mostly related to 1) the insecurity of not knowing what operations are being performed when other person is wearing the glasses and 2) the vulnerability derived from having a reduced view of the surroundings and few control over the collected data when the client is the one wearing them. It seems that there is a need for finding less obtrusive solutions for AR glasses as well as to make sure that users know and understand their capabilities and how their personal data is going to be treated. Additionally, some time is still required until wearing such devices in public is more socially acceptable, factor that seems to affect to a greater extent to people with an intuitive decision style. In the meantime, other ways for alleviating this issue could be explored, like more natural interaction methods that do not involve unusual gestures (as "air tapping" could be considered) or that help to better discern the wearer's intentions, so that other people around are more aware of them and less worried.

Some participants were also troubled about the trustworthiness of the recommendations and whether they could be biased towards the store's particular interests. Making clear the source of the provided information and its neutrality may be another relevant factor to increase the acceptance of applications like the one presented here.

What do these results mean for physical stores? Current AR technology is still young and well under development. Challenges in terms of social acceptance and privacy concerns can be expected to be overcome the more mature and available the technology becomes. As of today, however, it may be difficult to find a selling point to convince those users preoccupied for such matters, which stresses the importance of increasing customer awareness of privacy regulations and device capabilities. Nonetheless, AR is able already, at its current state, to provide a number of useful in-store functions, and there seems to be a great portion of consumers who are willing to try them. Stores may contribute in developing the potential of AR in physical retailing, while also making profit from it, by focusing on attracting these types of clients: those who enjoy shopping and those who find AR technology to be useful. That means developing applications and designing settings that contribute to an improved shopping experience and offer utilitarian benefits. Precisely this last part, focusing on the practical side of AR, may be a crucial factor in drawing the interest of those who have used AR in the past and no longer feel the so-called "wow" effect. Due to the current limitations of AR, it would be necessary to create specialized areas with different set-ups, more appropriate for experimenting and learning about products. In that regard, systems like the one here described may be more suitable for event-like shopping scenarios (such as trade fairs) where customer experience in a special setting is of the highest importance.

Limitations. Participants only watched a video of the system and were not able to test the prototype by themselves nor had access to the physical products; thus, they were not able to experience the "physical side" of the approach, perhaps the hardest part to imagine in conjunction with the other features of the system. Besides, participants were recruited through an online platform, which increases the probabilities for them to be more proficient in the use of technology and to have more trust on it than the average of the population. These factors may have had an impact on the assessment of the acceptance of the approach. There is also a probability of finding new types of customers if a bigger sample is used. In the same manner, the measurement of different psychological traits than the ones proposed here could uncover other still unknown relationships between them and the acceptance of the concept. Finally, few differences found between groups were actually statistically significant. However, considering the exploratory nature of this study, the observed results strongly suggest that such differences may indeed exist and offer a starting point for further research on the topic.

6 Conclusions

Related work on the topic of in-store AR acceptance has mainly focused on the independent assessment of technological, psychological and environmental factors. However, it is generally overlooked that connections between these aspects may exist that could be used to obtain a better understanding of how different types of clients accept the technology and what the obstacles are. To that end, this paper presents an exploratory study that investigates the interactions between a set of decision-making-related traits and their effects on the acceptance of an in-store AR advisor that runs on a head-mounted display.

Personal characteristics were defined by the Technology Adoption Propensity (TAP) [45], Chronic Shopping Orientation (CSO) [12] and Decision Styles (irrational, IDS, and rational, RDS) [22]. Using them to group participants uncovers the existence of four types of users within the sample, and a further analysis suggest differences between them in the acceptance of the system. There is an indication that technology-related aspects are not entirely responsible for defining AR acceptance, but that other factors are involved as well (and even negate the effects of high TAP scores). Overall results show that the approach is generally well-received, but users with low TAP scores are less convinced about its benefits and, thus, are less likely to make use of it. Both persons who know and trust technology (high TAP) and those who are experiential shoppers (high CSO) seem to be related to higher acceptance values. However, AR knowledge above the average seems to have a moderating effect on high TAP values. The most relevant issues to overcome involve concerns about privacy and social acceptance, opinion that is shared by all groups. Nevertheless, an intuitive decision style seems to be correlated with greater worries about the social acceptance of the approach.

The results indicate the existence of some psychological traits that have an impact on the acceptance of an AR-based in-store advisor and that relationships between them exist. Some of these relationships are revealed here, which may provide a deeper insight into how the concept could be introduced as a new in-store service by focusing on specific types of clients and making sure to cover privacy-related concerns. Nonetheless, further technology-dependant advances are still required in terms of intrusiveness for the approach to be more socially acceptable.

Future research should focus on discovering other underlying factors and on how to use them in both the improvement of the system to adapt to the specific needs of targeted consumer-types, and its successful introduction in real shopping environments.

References

1. Walgreens - aisle 411 and tango. http://aisle411.com/walgreens/ (2014) Accessed 19 May 2020
2. Olay Skin Advisor. https://www.olay.com/en-us/skinadvisor (2020). Accessed 22 May 2020
3. Ahn, J., Williamson, J., Gartrell, M., Han, R., Lv, Q., Mishra, S.: Supporting healthy grocery shopping via mobile augmented reality. ACM Trans. Multimedia Comput. Commun. Appl. **12**(1s), 16:1–16:24 (2015). https://doi.org/10.1145/2808207
4. Alavi, M., Joachimsthaler, E.A.: Revisiting DSS implementation research: a meta-analysis of the literature and suggestions for researchers. MIS Q. **16**(1), 95–116 (1992). https://doi.org/10.2307/249703, publisher: Management Information Systems Research Center, University of Minnesota
5. Álvarez Márquez, J.O., Ziegler, J.: Augmented-Reality-Enhanced Product Comparison in Physical Retailing. In: Proceedings of Mensch und Computer 2019, MuC 2019, Association for Computing Machinery, Hamburg, Germany, pp. 55–65 (September 2019). https://doi.org/10.1145/3340764.3340800
6. Álvarez Márquez, J.O., Ziegler, J.: In-store augmented reality-enabled product comparison and recommendation. In: Fourteenth ACM Conference on Recommender Systems, RecSys 2020, pp. 180–189. Association for Computing Machinery, New York (September 2020). https://doi.org/10.1145/3383313.3412266
7. Arth, C., Grasset, R., Gruber, L., Langlotz, T., Mulloni, A., Wagner, D.: The history of mobile augmented reality (2015)
8. Bagozzi, R.P.: The legacy of the technology acceptance model and a proposal for a paradigm shift. J. Assoc. Inf. Syst. **8**(4), 3 (2007)
9. Baier, D., Rese, A., Schreiber, S.: Analyzing online reviews to measure augmented reality acceptance at the point of sale: the case of Ikea. In: Successful Technological Integration for Competitive Advantage in Retail Settings, pp. 168–189. IGI Global (2015)
10. Beck, N., Rygl, D.: Categorization of multiple channel retailing in multi-, cross-, and omni-channel retailing for retailers and retailing. J. Retail. Consum. Serv. **27**, 170–178 (2015). https://doi.org/10.1016/j.jretconser.2015.08.001
11. Benbasat, I., Barki, H.: Quo vadis tam? J. Assoc. Inf. Syst. **8**(4), 7 (2007)
12. Büttner, O.B., Florack, A., Göritz, A.S.: Shopping orientation as a stable consumer disposition and its influence on consumers' evaluations of retailer communication. Eur. J. Market. (2014)
13. Buytaert, D.: Shopping with augmented reality. https://dri.es/shopping-with-augmented-reality (2018). Accessed 25 Jan 2021
14. Chatzopoulos, D., Bermejo, C., Huang, Z., Hui, P.: Mobile augmented reality survey: from where we are to where we go. IEEE Access **PP(99)**, 1–1 (2017)
15. Chylinski, M., De Ruyter, K., Sinha, A., Northey, G.: Augmented retail reality: Situated cognition for healthy food choices. In: ANZMAC (Australian and New Zealand Marketing Academy) (2014). 2017-09-11T01:54:50Z ISSN: 1447–3275

16. Cruz, E., et al.: An augmented reality application for improving shopping experience in large retail stores. Virtual Reality **23**(3), 281–291 (2019). https://doi.org/10.1007/s10055-018-0338-3
17. Dacko, S.G.: Enabling smart retail settings via mobile augmented reality shopping apps. Technol. Forecast. Soc. Change **124**, 243–256 (2017). https://doi.org/10.1016/j.techfore.2016.09.032
18. Dalim, C.S.C., Kolivand, H., Kadhim, H., Sunar, M.S., Billinghurst, M.: Factors influencing the acceptance of augmented reality in education: a review of the literature. J. Comput. Sci. **13**(11), 581–589 (2017)
19. Dillon, A.: User acceptance of information technology. Taylor and Francis, London (2001). 2006-07-20T00:00:01Z
20. Gentner, D., Medina, J.: Comparison and the development of cognition and language. Cogn. Stud. **4**(1), 1_112–1_149 (1997)
21. Gutiérrez, F., Htun, N.N., Charleer, S., De Croon, R., Verbert, K.: Designing Augmented Reality Applications for Personal Health Decision-Making (January 2019). https://doi.org/10.24251/HICSS.2019.212. 2019–01-02T23:56:24Z
22. Hamilton, K., Shih, S.I., Mohammed, S.: The development and validation of the rational and intuitive decision styles scale. J. Pers. Assess. **98**(5), 523–535 (2016)
23. Hilken, T., Heller, J., Chylinski, M., Keeling, D.I., Mahr, D., de Ruyter, K.: Making omnichannel an augmented reality: the current and future state of the art. J. Res. Interact. Market. **12**(4), 509–523 (2018). https://doi.org/10.1108/JRIM-01-2018-0023, publisher: Emerald Publishing Limited
24. Holdack, E., Lurie-Stoyanov, K., Fromme, H.F.: The role of perceived enjoyment and perceived informativeness in assessing the acceptance of AR wearables. J. Retail. Consum. Serv. 102259 (Septeber 2020). https://doi.org/10.1016/j.jretconser.2020.102259
25. Hopp, T., Gangadharbatla, H.: Novelty effects in augmented reality advertising environments: the influence of exposure time and self-efficacy. J. Curr. Issues Res. Advertising **37**(2), 113–130 (2016)
26. Huang, T.L., Liao, S.: A model of acceptance of augmented-reality interactive technology: the moderating role of cognitive innovativeness. Electron. Commer. Res. **15**(2), 269–295 (2015)
27. Huré, E., Picot-Coupey, K., Ackermann, C.L.: Understanding omni-channel shopping value: a mixed-method study. J. Retail. Consum. Serv. **39**, 314–330 (2017). https://doi.org/10.1016/j.jretconser.2017.08.011
28. Javornik, A.: Augmented reality: Research agenda for studying the impact of its media characteristics on consumer behaviour. J. Retail. Consum. Serv. **30**, 252–261 (2016). https://doi.org/10.1016/j.jretconser.2016.02.004
29. Kocas, C.: Evolution of prices in electronic markets under diffusion of price-comparison shopping. J. Manage. Inf. Syst. **19**(3), 99–119 (2002). https://doi.org/10.1080/07421222.2002.11045740
30. Kourouthanassis, P., Spinellis, D., Roussos, G., Giaglis, G.M.: Intelligent cokes and diapers: mygrocer ubiquitous computing environment. In: First International Mobile Business Conference, pp. 150–172 (2002)
31. Lackermair, G., Kailer, D., Kanmaz, K.: Importance of Online Product Reviews from a Consumer's Perspective (2013). https://doi.org/10.13189/aeb.2013.010101
32. Lancaster, K.J.: A new approach to consumer theory. J. Polit. Econ. **74**(2), 132–157 (1966)
33. Lee, Y., Kozar, K.A., Larsen, K.R.: The technology acceptance model: past, present, and future. Commun. Assoc. Inf. Syst. **12**(1), 50 (2003)

34. Linzbach, P., Inman, J.J., Nikolova, H.: E-commerce in a physical store: which retailing technologies add real value? NIM Market. Intell. Rev. **11**(1), 42–47 (2019). https://doi.org/10.2478/nimmir-2019-0007, publisher: Sciendo Section: NIM Marketing Intelligence Review
35. Ludwig, T., Hoffmann, S., Jasche, F., Ruhrmann, M.: Vacuumcleanar: augmented reality-based self-explanatory physical artifacts. In: Proceedings of the Conference on Mensch und Computer, pp. 291–302 (2020)
36. Masood, T., Egger, J.: Augmented reality in support of industry 4.0–implementation challenges and success factors. Robot. Comput. Integr. Manuf. **58**, 181–195 (2019). https://doi.org/10.1016/j.rcim.2019.02.003
37. McLean, G., Wilson, A.: Shopping in the digital world: Examining customer engagement through augmented reality mobile applications. Comput. Hum. Behav. **101**, 210–224 (2019). https://doi.org/10.1016/j.chb.2019.07.002
38. Narayana, C.L.: Graphic positioning scale: an economical instrument for surveys. J. Market. Res. **14**(1), 118–122 (1977)
39. Pantano, E.: Innovation drivers in retail industry. Int. J. Inf. Manage. **34**(3), 344–350 (2014). https://doi.org/10.1016/j.ijinfomgt.2014.03.002
40. Parasuraman, A.: Technology readiness index (tri) a multiple-item scale to measure readiness to embrace new technologies. J. Serv. Res. **2**(4), 307–320 (2000)
41. Park, Y., Gretzel, U.: Influence of consumers' online decision-making style on comparison shopping proneness and perceived usefulness of comparison shopping tools (2010)
42. Perannagari, K.T., Chakrabarti, S.: Factors influencing acceptance of augmented reality in retail: insights from thematic analysis. Int. J. Retail. Distrib. Manage. **48**(1), 18–34 (2019). https://doi.org/10.1108/IJRDM-02-2019-0063, publisher: Emerald Publishing Limited
43. Poushneh, A., Vasquez-Parraga, A.Z.: Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. J. Retail. Consum. Serv. **34**, 229–234 (2017). https://doi.org/10.1016/j.jretconser.2016.10.005
44. Rashid, Z., Pous, R., Melià-Seguí, J., Morenza-Cinos, M.: Mobile augmented reality for browsing physical spaces. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp 2014 Adjunct, pp. 155–158, New York, ACM (2014). https://doi.org/10.1145/2638728.2638796, event-place: Seattle, Washington
45. Ratchford, M., Barnhart, M.: Development and validation of the technology adoption propensity (tap) index. J. Bus. Res. **65**(8), 1209–1215 (2012)
46. Rauschnabel, P.A., He, J., Ro, Y.K.: Antecedents to the adoption of augmented reality smart glasses: a closer look at privacy risks. J. Bus. Res. **92**, 374–384 (2018)
47. von Reischach, F., Guinard, D., Michahelles, F., Fleisch, E.: A mobile product recommendation system interacting with tagged products. In: 2009 IEEE International Conference on Pervasive Computing and Communications. IEEE (Mar 2009). https://doi.org/10.1109/percom.2009.4912751
48. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
49. Roy, S.K., Balaji, M.S., Quazi, A., Quaddus, M.: Predictors of customer acceptance of and resistance to smart technologies in the retail sector. J. Retail. Consum. Serv. **42**, 147–160 (2018). https://doi.org/10.1016/j.jretconser.2018.02.005
50. Schafer, J.B., Konstan, J., Riedi, J.: Recommender systems in e-commerce. In: Proceedings of the 1st ACM conference on Electronic commerce - EC 99. ACM Press (1999). https://doi.org/10.1145/336992.337035

51. Scott, S.G., Bruce, R.A.: Decision-making style: the development and assessment of a new measure. Educ. Psychol. Measur. **55**(5), 818–831 (1995)
52. Spreer, P., Kallweit, K.: Augmented reality in retail: assessing the acceptance and potential for multimedia product presentation at the pos. Trans. Market. Res. **1**(1), 20–35 (2014)
53. Sprotles, G.B., Kendall, E.L.: A methodology for profiling consumers' decision-making styles. J. Consum. Aff. **20**(2), 267–279 (1986)
54. Välkkynen, P., Boyer, A., Urhemaa, T., Nieminen, R.: Mobile augmented reality for retail environments. In: Proceedings of Workshop on Mobile Interaction in Retail Environments in Conjunction with MobileHCI (2011)
55. Verhoef, P.C., Kannan, P., Inman, J.J.: From multi-channel retailing to omni-channel retailing. J. Retail. **91**(2), 174–181 (2015). https://doi.org/10.1016/j.jretai.2015.02.005
56. Wassom, B.: Augmented Reality Law, Privacy, and Ethics: Law, Society, and Emerging AR Technologies. Syngress, Waltham (2014)
57. Webster, J., Trevino, L.K., Ryan, L.: The dimensionality and correlates of flow in human-computer interactions. Comput. Hum. Behav. **9**(4), 411–426 (1993). https://doi.org/10.1016/0747-5632(93)90032-N
58. Westbrook, R.A., Black, W.C.: A motivation-based shopper typology. J. Retail. **61**(1), 78–103 (1985)
59. Zhu, W., Owen, C.B.: Design of the promopad: an automated augmented-reality shopping assistant. JOEUC **20**, 41–56 (2008)

Paper VIII. Creating Omni-Channel In-Store Shopping Experiences through Augmented-Reality-Based Product Recommending and Comparison

Reproduced with permission from the Taylor & Francis Group: Jesús Omar Álvarez Márquez and Jürgen Ziegler. *Creating Omni-Channel In-Store Shopping Experiences through Augmented-Reality-Based Product Recommending and Comparison*. International Journal of Human–Computer Interaction, 1-26.

https://doi.org/10.1080/10447318.2022.2163650

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2163650

Taylor & Francis
Taylor & Francis Group

Creating Omni-Channel In-Store Shopping Experiences through Augmented-Reality-Based Product Recommending and Comparison

Jesús Omar Álvarez Márquez and Jürgen Ziegler

Interactive Systems Research Group, University of Duisburg-Essen, Duisburg, Germany

ABSTRACT

We present a novel approach to the inclusion of online services within a physical retailing scenario through the use of augmented reality (AR) technology and head-mounted displays (HMD). A concept has been designed for providing the innovative combination of product data, comparison support and product recommendations, all directly accessible from the pertinent physical products. Prototypes were implemented and evaluated in several user studies to address questions related to information acquisition, product exploration and user acceptance, including a comparison against a baseline system (i.e., smartphone app without AR capabilities). The results indicate that providing in-store functions *via* AR HMDs can be on par with a non-AR smartphone approach in terms of practicality, and may provide superior benefits concerning the discovery and exploration of products, and the perception of their differences. However, effort must be dedicated to design AR UIs able to compete against the clearer and more structured information visualization of traditional displays.

1. Introduction

In recent years, augmented reality (AR) has become a well-established technology that has been successfully applied in many fields, such as industry, entertainment, medicine and education (de Souza Cardoso et al., 2020; Parekh et al., 2020; Wu et al., 2013). Currently, the technology is growing in popularity among the general audience thanks to the commitment of leading companies, such as Meta, Microsoft, or Apple, to its further development, also fostering the emergence of new terms like Extended Reality (XR) or Metaverse (Mystakidis, 2022).

Thanks to the new opportunities brought by AR in terms of information access and interaction (Kim et al., 2018), the technology has also gained interest in the area of physical retailing (Caboni & Hagberg, 2019), as it offers customers a new and efficient form of delivering detailed product information on the spot without having to consult shop personnel, and can provide new shopping experiences, for example, by highlighting relevant features in situ, or by explaining product use (Ludwig et al., 2020; Zimmermann et al., 2022), all of which enhances the customer experience. However, and even with the added value of richer product information provided by AR functions, physical shops today are only one of the customer touchpoints where purchases are made. In that regard, the massive increase in online shopping sites and their advantages over physical stores (Perea y Monsuwé et al., 2004) did indeed pose a threat to traditional shops in the past, but more recent predictions indicate that this may not be the case anymore (Doherty & Ellis-Chadwick, 2010). Instead, new trends are more inclined toward the implementation of omni-channel solutions where physical and online outlet channels are intertwined (Huré et al., 2017; Verhoef et al., 2015). Despite this, they are still far from a strong integration, often limited to online ordering and physical pick-up, or the contrary, using physical stores as show-rooms preceding an online purchase (Gao & Su, 2019). AR offers possibilities for a much stronger integration of online and physical channels (Hilken et al., 2018), but these possibilities are still under-researched and, if they exist, limited (Riar et al., 2022). One can observe the following gaps in the current use of AR in physical shops which we address in our research:

- First, the in-store application of AR can go beyond the conventional provision of product information through AR visualizations, delivered either by smartphone or, still rarely, through special AR devices, by better connecting the products present in the shop with online offers. Since shop space is expensive, the selection of products on display is necessarily limited thus constraining users' choice and possibly making them uncertain about whether the best option is really available. Including online items in an AR presentation extends the space of products that can be explored which may give users the impression of a more comprehensive and fair offering.
- Second, by providing users with the possibility to compare in-store products with online offers on a feature

CONTACT Jesús Omar Álvarez Márquez jesom.alma@gmail.com Interactive Systems Research Group, University of Duisburg-Essen, Duisburg, Germany

level allows users to decide which offer is better suited for their purpose. The spatially referenced display of features on the in-store products helps users understand the features' meaning and relevance even for a remote product. Side by side comparison of in-store and remote products also supports assessing the visual and aesthetic quality of the product.

- A third type of functionality that has been mostly overlooked in AR research for in-store use is recommender functionality. Recommender systems are a well-established technology in online shops but as yet only few examples exist of its combination with physical shopping environments. Recommender systems research has developed a large range of preference elicitation and recommending methods (Ricci et al., 2011) but limited to online-only settings. Yet, seeing and touching, possible operating, an in-store product can play an important role in the user's preference formation since features and appearance of the physical product may act as cues for becoming aware of one's actual needs and preferences (Grohmann et al., 2007). An in-store product the users focusses upon can thus serve as an anchor through which the system can recommended similar products even if they are only available online.
- Finally, there is also only limited knowledge concerning users' psychological characteristics that may influence their assessment of in-store AR and their acceptance of the technology, more so when considering its utilization for pragmatic rather than entertaining or advertising purposes.

Based on these gaps in current research, we formulate the following research questions:

RQ1: How can product recommendation and comparison functions be delivered through AR in a hybrid physical-online setting?

RQ2: What are the advantages of providing AR-based functions in a physical shopping scenario during the decision-making process in relation to information access, product discovery and purchase confidence?

RQ3: What are the implications that this hybrid physical-digital setting may have on the way users learn and explore the digital space?

RQ4: What are the effects that user psychological traits have on the acceptance of such systems, and to what extent does the use of AR head-mounted displays influence the acceptance of in-store support functions among consumers with different psychological profiles?

Addressing these research questions, this article presents a novel approach for amplifying the in-store shopping experience, together with its implementation into a fully working prototype for AR head-mounted displays (HMD), and its evaluation in a series of user studies. The system is capable of detecting physical products and providing superimposed, contextual information *via* augmented reality, which can be directly compared against that of the other products. Furthermore, recommendations of similar products are generated for each item at the store. These recommendations may contain products not physically available, which in turn extends the stores' product catalog and supports the union of both digital and physical retailing channels.

Over the multiple stages of this research (Álvarez Márquez & Ziegler, 2019, 2020, 2021), the described concept and developed system were evaluated with a focus on the possible implications of providing AR functions in physical settings in terms of information access and purchase decision-making, while also considering the effects of the presence of physical items in the decision-making process and the role of user psychological traits in the acceptance of the approach. This article integrates these findings and makes a further contribution by investigating the acceptance of the system when compared to a more traditional take on in-store digital services. To that purpose, a smartphone app with equivalent functionality was implemented and used as baseline in two complementary user studies: an exploratory, online study to obtain an initial overview of how the systems were perceived regarding a broader set of constructs, followed by a more focused lab study to gather first-hand information on usability and user experience. To the best of our knowledge, no previous study exists that makes such comparison, the results of which may provide valuable insight into the real implications of incorporating AR based functionality in physical stores though dedicated AR devices, in contrast to doing it with a more standard technological approach.

The background section discusses the current status of physical retailing and the new possibilities brought by AR, and presents the possible advantages of providing in-store shopping support. Section 3 describes the characteristics of the developed prototype for AR head-mounted displays. Section 4 gives an overview of previous evaluation studies, which were of formative nature and allowed an incremental user-centric development of the prototype. Section 5 presents the results of an online study focused on the impact of psychological user characteristics on the acceptance of AR-based in-store shopping support functions, including a comparison of the developed system against a non-AR-based approach for smartphones. Section 6, on the other hand, presents the results of an interactive lab evaluation where such functions were tested for both AR and non-AR based approaches, with an emphasis on user experience and usability aspects. Final conclusions and further work are presented in Section 7.

2. Background

2.1. Developments in omni-channel retailing

Online shopping is often preferred over physical stores due to its advantages in terms of information accessibility, expanded product selection, and ease of comparison (Perea y Monsuwé et al., 2004), and it offers a plethora of solutions oriented to support the client's decision making process, such as comparison tools (Kocas, 2002; Park & Gretzel, 2010), price trackers, customer reviews and ratings (Lackermair et al., 2013), detailed descriptions or product recommendations (Schafer et al., 1999). On the contrary, obtaining relevant information in physical settings often requires interacting with the store's staff (Homburg et al., 2011), or reading it from fliers or posters, where it is often challenging to find detailed and reliable information. As a consequence, clients often resort to external sources (e.g., specialized online portals) or make uninformed purchase decisions. Nonetheless, physical retailing still has benefits hardly replicable in online settings. This is especially true for those customers who value its experiential and social aspects, and those who rely on physical interaction with products for making a purchase decision, as these are the main strengths of brick-and-mortar stores (Enders & Jelassi, 2000).

Due to the pros and cons of each shopping medium, it has become clear that neither will be able to completely outperform the other for the time being (Doherty & Ellis-Chadwick, 2010). Therefore, companies have adopted a more conciliatory attitude toward traditional retailing and e-commerce, as shown by how common it has become to offer online channels that work in parallel or in combination to the physical ones. Different levels of integration between the available channels have been identified, named as multi-channel, cross-channel or omni-channel (Beck & Rygl, 2015; Jasin et al., 2019). Multi-channel retailing is the most widespread model (in which each channel works independently), but the omni-channel approach has started to become popular and, at the moment, it seems a safe prediction of how the future of commerce will look like (Huré et al., 2017; Verhoef et al., 2015). Omni-channel retailing stands for the greatest level of channel integration, consisting on the elimination of the boundaries between physical and virtual mediums (from a consumer's perspective) for a more rewarding shopping experience. This new point of view opens a door to the creation of novel shopping concepts, where the features of digital environments could be brought into physical settings.

2.2. AR for consumer applications

People's interest in AR has increased steadily in recent years (McCluskey, 2022), along with the investment that firms make on the integration of AR solutions within their commercial activities, either as products or production tools (Rese et al., 2017; Roitman et al., 2017). However, although AR has found its place in leisure and advertising domains (Chatzopoulos et al., 2017), as well as in training and education fields, it hardly has an impact on the life of the average consumer. Part of the issue is related to how hard is to retain the user's interest when the sense of novelty (a quality often exploited with advertising purposes) fades away (Hopp & Gangadharbatla, 2016), which is aggravated by the existence of more traditional approaches that may achieve similar results, and therefore are chosen over an AR solution (Chatzopoulos et al., 2017). In this regard, the acceptance of AR-based applications greatly relies on its perceived ease of use and perceived usefulness (Huang & Liao, 2015), which indicates how important it is to design intuitive methods to interact with AR-UIs and apply them to scenarios where its practical value outweighs that of other approaches.

AR solutions have shown to be of value within the field of retailing (Caboni & Hagberg, 2019), and can be beneficial at various consumer touch points thanks to the growth of mobile technology (Javornik, 2016), and mobile AR apps have shown to enhance the shopping experience (Dacko, 2017). More specifically, AR has been pointed as having a positive impact on customer engagement, customer satisfaction and purchase intention (Bonetti et al., 2019; Pantano, 2014; Poushneh & Vasquez-Parraga, 2017). Moreover, AR has the additional advantage of being an enabler of omni-channel experiences by allowing the seamless integration of digital and physical retailing channels (Beck & Rygl, 2015; Dacko, 2017; Hilken et al., 2018). In spite of this, omni-channel experiences enabled *via* AR are rarely studied for in-store scenarios, while most current research interest is oriented toward the use of mobile devices and try-at-home applications in online settings (Riar et al., 2022). Nonetheless, the field of AR technology is still heavily under development, and new improvements on dedicated hardware may make it more accessible and realistically usable in physical retailing scenarios, and risks assumed by companies, such as the need of an initial investment for its implementation and staff training (Dacko, 2017) may as well decrease over time.

2.3. Providing in-store AR-based shopping support

Consumers often lack well-defined preferences, and tend to build them on the spot (Payne et al., 1992), which in conjunction to other issues like choice overload (Bollen et al., 2010), may pose an obstacle for making a satisfactory purchase decision. To overcome these limitations, shopping advisors offer a variety of useful tools aimed at helping users to make a more confident purchase, including, in many cases, recommending capabilities (Schafer et al., 1999). Within the domain of AR, previous research has found that the technology can improve the search of information at the point of sale (Spreer & Kallweit, 2014) and support clients in making a purchase decision (Chylinski et al., 2014), which indicates its potential for developing in-store shopping assistants (Mora et al., 2020; Zimmermann et al., 2022). More particularly, AR technology can be used to support clients in a number of ways, including: effortless acquisition of product information; contextual explanations that help users learn about the product space; improved product discovery and preference construction thanks to the integration of recommending techniques; expanded product catalog that joins online and physical retailing channels; and product comparison aids to facilitate the customer's purchasing choice. In line with all these potential benefits, the following subsections address features that are commonly seen in modern shopping assistants, and pay particular attention to the possibilities of implementing them *via* AR.

2.3.1. Extending product information through AR

One of the most critical downsides of physical stores when compared against digital ones has to do with the acquisition of information (Perea y Monsuwé et al., 2004). This issue can be aggravated if the product domain is not well known by the client, more so within the spectrum of technical items that require expert knowledge to understand their attributes. It is also important to consider that a shopping context involves decision-making, and since processing information has a cost, consumers can feel frustrated when information is either insufficient or excessive, as it increases the complexity of the decision (Wierenga & Van der Lans, 2008).

AR can be a viable solution to provide easy access to on-site, reliable information, as has been demonstrated by similar research on the tourism domain (Kounavis et al., 2012). In the retailing field, examples of providing extended product information *via* AR include the work by Välkkynen et al. (2011), where augmentations are used to present the content of closed packages; Gutiérrez et al. (2019) developed a system that delivers on-site health-related product information, and investigate different visualization layouts and their suitability for different AR platforms; Cruz et al. (2019) explored the possibility of providing navigation support for guiding clients in large retail stores; and Ludwig et al. (2020) studied the benefits of using AR to expose the technical features of physical products.

As previous research has shown, AR can be used to acquire in-store information in unprecedented manners, supported by the contextual awareness and information readiness enabled by the technology. However, although some benefits of its use have already been outlined, it is still unclear to what extent AR is actually advantageous compared to other approaches (such as a specialized non-AR smartphone app). Further research is also needed to investigate whether the physicality of products is important to the point that providing in-store information access contributes to the decision-making process, in contrast to conducting the pre-purchase search of information by other means (e.g., by researching websites at home).

2.3.2. Providing product search and recommendations through AR

Recommender systems (RS) play a key role in reducing the amount of information that consumers need to evaluate by providing users with suitable purchasing alternatives. They have proven usefulness that is backed by extensive research (Ricci et al., 2011). RS have therefore become an everyday tool for most internet users. In addition, RS provide a number of benefits to online shops by increasing the likelihood of visitors to become buyers, as well as cross-selling opportunities and improved consumer loyalty (Schafer et al., 1999). As common as RS are in digital scenarios, they have rarely been implemented in physical ones, despite their potential to be equally valuable when dealing with physical objects. For instance, Kourouthanassis et al. (2002) presented an in-store shopping advisor to provide recommending features in smartphones: a system that uses RFID technology for detecting products to automatically create and update a shopping list, while also offering product information and recommending personalized promotions. Another example is found in APriori (von Reischach et al., 2009), a system for mobile devices that provides product data, recommendations and user ratings; similarly, RFID technology was used by Chen et al. (2015) to design a smart shopping environment in which clients receive recommendations according to their purchase history; or Fagerstrøm et al. (2020), who explored the benefits of offering personalized recommendations by using the Internet of Things (IoT) technology. Nevertheless, none of these in-store recommendation approaches considers the use of AR.

AR technology brings new opportunities to present recommendations and the real world objects they concern together in a shared space. This union may offer pragmatic benefits in terms of search and decision support (Walter et al., 2012) while also creating an engaging experience (Pantano, 2014). This idea is supported by the results of previous work: Ahn et al. (2015) assessed, among other aspects, the advantages of using AR for product search in retail stores; Acquia Labs (Buytaert, 2018) designed a system capable of offering product information and in-store navigation support; although focused on in-home services, the results obtained by Huynh et al. (2018) suggest that AR-enhanced, personalized product recommendations offer some benefits in contrast to browser based UIs; Torres-Ruiz et al. (2020) combined the IoT and AR technologies to offer recommended itineraries within a museum based on the user's interest; Mora et al. (2020) discussed the requirements of in-store assistants and their implementation using mixed reality, which also included tailored product recommendations; and Zimmermann et al. (2022), whose investigation highlights the benefits of in-store AR assistance in terms of usefulness, entertainment and informativeness in contrast to unassisted shopping. The approach has been successfully used commercially too, as demonstrated by the partnership between Aisle411 and Tango in the creation of an app for Walgreens stores (Aisle 411 & Tango, 2014) to deliver product information, promotions and navigation; or the Olay Skin Advisor (2020), which detects the consumer's face skin condition and recommends suitable products.

The existent research reveals a growing interest in AR as a medium through which to provide recommendations, both in scientific and commercial contexts. Nonetheless, current research has not yet explored the opportunity of including digital-only products within the range of possible recommendations, nor the implications that such union may have in how users browse and learn about digital and physical spaces. Moreover, techniques for influencing the outcome of the recommendations have rarely been included in physical contexts, such as attribute critiquing of products, which, together with direct product inspection, may have an impact on how clients of a store create and progressively develop their preference models.

2.3.3. Supporting product comparison through AR

Comparison plays a significant role in how humans understand, discover and evaluate their surroundings (Gentner &

Medina, 1997). Literature in the field of consumer behavior also highlights its importance when making a purchase decision, a situation where consumers tend to focus on the attributes of the different options and compare them against each other (Lancaster, 1966). However, comparison capability is constrained by the number of attributes to be compared and how accessible they are due to the limitations of short-term memory (Alvarez & Cavanagh, 2004). Online retailers address this issue by offering comparison tools (Kocas, 2002; Park & Gretzel, 2010), but this type of services is difficult to implement when the items to compare are part of a physical setting.

On this topic, AR can alleviate the mental workload of retaining information by using spatial superimposition (Tang et al., 2003), and has been highlighted as a suitable medium with which to support product comparison in physical retailing (Mora et al., 2020). Another potential advantage appears when AR is used for both presenting digital recommendations and providing comparison features: consumers may have then the possibility to better appreciate the physical qualities of digital products (e.g., size) when these are presented next to real items in the store. In relation to this idea, although in a different direction, previous research supports that AR is able to increase the tangibility of digital elements (Hilken et al., 2017; Overmars & Poels, 2015; Verhagen et al., 2014). Also related is the research focused on "discrepancy check" of physical items and their digital counterparts, as shown by Georgel et al. (2007), where real world construction sites are compared against their original 3D blueprints. Nonetheless, the approach here presented of hybrid (digital-physical) product comparison remains unique within its field of application. Therefore, open questions appear considering this uncommon scenario: firstly, with regard to the general effects that offering comparison support applied to a physical context has on the purchase decision; and secondly, the extent to which such hybrid setting supports the understanding of the characteristics of digital items when these are learned by taking the physical ones as reference.

2.4. Acceptance of AR technology in retailing contexts

In general, technology acceptance, or as it has been defined: "*the demonstrable willingness within a user group to employ IT for the tasks it is designed to support*" (Dillon, 2001), seems to greatly rely on factors, such as perceived ease of use (PEOU), perceived usefulness (PU), or perceive adaptiveness (Roy et al., 2018). Precisely, one of the most widely used approaches for acceptance assessment, the Technology Acceptance Model (TAM) (Lee et al., 2003), considers PU and PEOU as the main drivers of technology adoption. This appears to be true for AR applications in retailing contexts too, as TAM is also the most utilized acceptance measurement tool in this case (Perannagari & Chakrabarti, 2019). Therefore, the significance of finding utilitarian uses for AR and designing intuitive interfaces and interaction methods is further highlighted when considering the role that these elements play in the acceptance of the technology (Rauschnabel et al., 2018). Apart from these factors, previous research indicates that the adoption of AR is also dependent of security and privacy aspects, as its usage is often perceived to pose risks in that respect (Rauschnabel et al., 2018; Wassom, 2014).

Nonetheless, care must be taken when regarding the significance of these factors as absolute, and attention must be paid to the existence of consumer types and how they react to the technology. Customers with different psychological profiles may respond in different ways to the proposed AR-based in-store services, especially in the case of this kind of innovative (and perhaps conspicuous, if using HMDs) approaches. Merely assessing the technology acceptance of a user (via TAM or any other mean) may not be sufficient to fully determine the drivers behind the acceptance or rejection of AR in these type of complex scenarios, which calls for the study of other relevant factors, like the decision or shopping styles of the client. Per contra, questions concerning the existence of consumer types, and the study of their acceptance of AR in retailing settings, have been generally overlooked in previous research.

All things considered, there are sufficient reasons to believe that the use of in-store assistants may improve the overall shopping experience by providing clients with useful, personalized information, while also allowing them to test products themselves. Furthermore, AR technology appears to be particularly well suited to be the platform through which to provide these types of in-store services. Nonetheless, while an AR approach has already been proven beneficial and commercially successful in the shape of "magic-mirrors" (or virtual try-on) (Beck & Crié, 2018; Javornik et al., 2016; Kim & Forsythe, 2008; Smink et al., 2019), the combination of offering product information, customized recommendations and hybrid product comparison *via* AR in a physical retailing scenario still is a mostly unexplored area. Due to the unusual characteristics of the concept, research gaps exist regarding each one of the mentioned functions individually, but also generally in terms of the user acceptance of this kind of AR-based services.

3. AR-based in-store shopping support prototype

We propose an approach to support clients of physical stores by providing access to services normally only available in online platforms and that appear seamlessly integrated within the physical shopping environment thanks to AR technology, while customers are still able to perform direct product inspection. Products in the store are automatically detected by the device, and contextual information (i.e., relative to each product) is displayed next to them, including product recommendations not only based on attribute similarity, but also on visual likeness. Finally, these recommendations can be further refined by using critiquing functions on the attributes of the physical product on which they are based. The approach is most appropriate for products with a relatively complex and broad set of attributes, provided that their physical characteristics also play a role in the purchase decision (otherwise, the physical presence of the product would be irrelevant).

The approach is designed for head-mounted displays (HMD), and implemented specifically to run under Microsoft's HoloLens. A HMD approach was chosen over a smartphone one because it offers more interesting interaction possibilities, as it allows for hands-free inspection of products and gaze-based selection, which eliminates the constraint of using one hand to hold the device. Moreover, HMDs enable a more natural and integrated visualization of digital augmentations by presenting them directly superimposed over the real world, in contrast to the disconnection between worlds resultant of watching them on the screen of a smartphone. Altogether, HMD technology better represents the paradigm of future AR, it is currently under heavy development, and its relevance and availability will most likely grow in following years. The concept may be suitable for a broad set of product domains, but in this particular case vacuum cleaners, which are common and technical commodities, have been chosen.

In the following, the functionality of the implemented prototype is presented (Figure 1), with a focus on the three main functions of *information acquisition*, *delivery of recommendations*, and *product comparison*, all of them aiming to the integration of physically present products with ones only available online.

Figure 1. Main view of the system. Attribute categories and recommendations are placed to the left and right of the product, respectively.

3.1. Access to product information

The developed prototype is able to automatically detect the position and orientation of physical products by using marker recognition (thanks to the Vuforia Engine[1]) and present context-based information relative to them. Because object recognition is not reliable enough yet, markers placed on the front of the products were used instead, but were designed to be as unobtrusive as possible by matching the product's own style/colors. When a user has visual focus on a certain product, digital augmentations with relevant data are displayed surrounding the item (left side of Figure 1). The displayed information can be easily switched from one product to another by simply directing the visual focus to the desired one. Users can interact with most of the provided data by using the "air tap" gesture, which usually discloses more detailed information. Product information is organized as follows:

Categories: product attributes are grouped in categories to prevent overcrowding the limited visual space. At the same time, categories have a score that summarizes how well the product performs in each one of them, which provides a quick overview of its qualities. "Air tapping" on a category discloses the attributes within.

Attributes: the system shows the values of each one of the attributes of the product (Figure 2). Attributes can be selected like categories are (by "tapping" on them), which uncovers extra information (a brief description and its location on the product, when applicable—Figure 3) and discloses new options for bookmarking and critiquing them (more about this in Section 3.2). Some attributes may contain sub-attributes (e.g., a type of filter has a replacement time that varies between products), in which case they are also revealed after the selection is performed.

The system includes short-cuts to quickly swap from one category to another, and a feature for bookmarking attributes, which permits the creation of a custom category where only user-selected attributes are located.

Although providing centralized information and showing it conveniently next to a product are good points in favor of AR, the fact is that similar results could be achieved by more standard means (e.g., a smartphone app with object recognition capabilities). However, by including relevant spatial information for each attribute (i.e., their physical location), AR adds a new layer of interaction between digital and physical worlds, and their connection is made more apparent. This union may not only encourage users to physically explore products, but also contribute to a better understanding of their features.

3.2. Product recommendations

The system provides product recommendations based on the product the user is currently looking at, for which four recommended items are displayed at a time (Figure 4). These items are generated by using a content-based

Figure 2. Attributes within a category.

Figure 3. Tapping on an attribute shows its location on the product and a brief description.

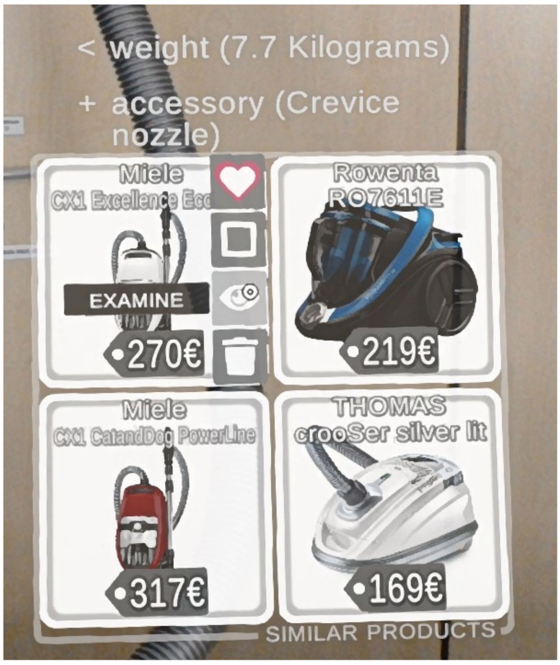


Figure 4. Recommended items and critiqued attributes (above them).

technique that takes into account the attributes of the particular product they are attached to, which are used as the user's initial preference model. This means that browsing other physical products will bring different recommendations, allowing users to explore and filter the product space by focusing on the items that are most relevant to them.

Product similarity scores are calculated considering not only the technical characteristics of the products, but also their visual appearance. For the assessment of the latter, images of all products are compared by using DeepAI's API,[2] which returns a value indicating image similarity (0 for completely identical ones). This value is not calculated at runtime; instead, for each available pair of products, their image comparison value is stored in the database beforehand to improve the recommender's response time. When multiple images of a product are available, only the lowest returned value is stored. The final outcome of the recommending algorithm consists of a list of products ordered by similarity score, of which the top four are displayed to the user as recommendations.

Users can influence the outcome of the recommendations by either removing those that they do not like (which shows the next most similar ones) or, as another innovative feature, by critiquing specific attributes of the base physical product. Using the latter option requires clients to choose an attribute and use the critiquing buttons (disclosed next to it) to select whether they seek for recommendations with higher or lower values than the present product for that particular property. For non-numerical attributes, clients must choose whether they should be included or excluded from the recommendations instead. By performing the critiquing action, the user's preference model is modified, the recommendations are refreshed accordingly and new items are provided. In the most recent version of the prototype, users can customize the value of a critiqued attribute directly after its definition, which gives a higher degree of control over the outcome of the recommending process.

After the selection of a digital recommendation, if it represents a product that is physically available at the store, the user can find it by following a virtual compass that marks its location. However, the system allows the inclusion of recommendations that do not possess a physical counterpart, that is, products that are not physically available at the store but that belong to the retailer's online catalogue. This is still a widely under-explored aspect that may hold great value from a commercial standpoint: the inclusion of recommendations brings the opportunity to expand the store's catalog by recommending products not physically available, but that consumers may discover and experience by examining physical ones with similar characteristics, an effect that can be further amplified by the inclusion of comparison support.

3.3. Product comparison support

Due to the importance that comparing holds during the purchase decision-making process, and taking online stores as reference, on-site product comparison support has been a cornerstone element in the development of this approach to AR-based shopping advisors.

In the prototype, users can select up to three products by "tapping" on them (no matter whether these are physical or digital-only items), which in turn initializes the comparison view. In this mode, when a user looks at one of the selected products, the attributes of the out-of-sight ones are displayed side-by-side to those of the former (Figure 5), together with the following visual aids:

Color coded attributes: A color is assigned to each product, which is used to code its attributes and make them more easily distinguishable.

Performance indicator: The attributes of a product are highlighted in green or red based on how well (or bad) it performs in each particular aspect relative to the other selected items.

Visual superimposition: Some attributes are suitable for a more "visual" comparison. In these cases, the system superimposes the attributes of the out-of-sight products over the currently seen one (e.g., scales displaying their size).

The system also has a function to "save" products and store them in a virtual shopping cart. This cart follows the client and grants direct access to the items placed in it. Saved products can be selected again at any moment for comparison purposes, without the need for the user to go to their physical location each time. Recommendations can be selected and saved the same way as physical products are, which makes possible to compare digital-only products against physical ones.

4. User-centric development of the prototype

The design and implementation of the prototype followed a user-centric incremental approach, by which the user interface was improved and specific functions added in several iterations, based on the feedback obtained in user studies. This type of system development process helped to bring insight into how in-store AR functions provided *via* dedicated hardware can be more effectively implemented (RQ1). The outcome of each phase resulted in a functional system, which allowed for an early evaluation of particular features and their related research questions in a series of consecutive studies. All studies followed the ethical guidelines of our local ethics committee, and all participants signed informed consent forms, whereby they agreed to the use of the collected data for research purposes.

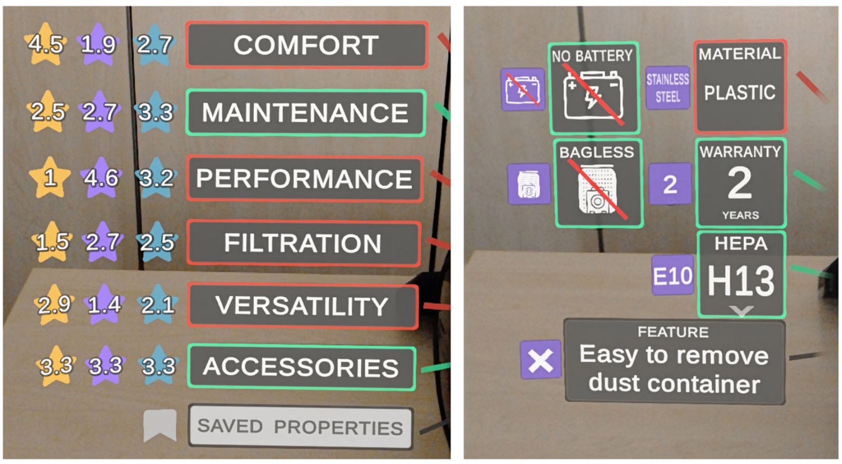


Figure 5. Comparison view for categories (three products are compared in the example) and attributes (two products). Colors indicate to which product they belong to. When comparing attributes (right image) only the properties of the products that are not being directly looked at are color-coded. Red and green frames indicate for which attributes/categories the current product has worse and better values than the other items in the comparison. (a) Product information. "Overview" category is shown (scores per category). (b) Recommendations. Base product and some critiqued attributes have been defined. (c) Comparison view. Attributes in the "Performance" category are shown.

Because these studies were originally covered in previous publications, this section focuses on providing a broader view of their individual outcomes and how their findings relate to each other. The particular setting and scope of each previous study are summarized in the following lines, after which their results are described in more detail.

Development iteration 1 (Álvarez Márquez & Ziegler, 2019): The initial prototype had functions to offer contextual product information and support the comparison of up to three physical items. It was evaluated in a lab study, with a total of 50 participants (38 female, average age of 21.16, σ 3.525, all of them German residents and mostly students with a technological background) and three physical vacuum cleaners as available products. The goal was to assess different interaction and visualization methods, and to obtain an early view of the possible benefits of using in-store AR-based comparison functions. It was also in the scope of the study to assess the importance of physical presence of products in the purchase decision, in order to make a first judgment on the advantages of combining digital assistance and direct product inspection within a physical retailing context.

Development iteration 2 (Álvarez Márquez & Ziegler, 2020): The user interface of the prototype was completely remade to improve its usability following the feedback gathered during the evaluation of the previous iteration. Recommending functions were also implemented, through which four items similar to the one the user was focused on were displayed, and could be further refined *via* attribute critiquing. The system used two different similarity scores for generating recommendations: two of the recommended items were based only on technical attributes, while the other two considered only their visual appearance (by examining visual-related attributes and comparing images of the items). The inclusion of recommendations also allowed the incorporation of a digital catalog that could be accessed through them.

This smaller lab study had 10 participants (four female, average age of 28.1, σ 4.06, all of them German residents and mostly students with a technological background), three physical vacuum cleaners, and 97 extra ones from the digital catalog (for a total of 100). The main goal was to further investigate the effects that a digital-physical shopping scenario may have in the definition of user preferences, that is, how users learn about product characteristics when these can be directly inspected on the real product; how users perceive and browse the digital space when the physical one is taken as reference (also considering possible anchoring effects); and how the overall setting influences purchase confidence. The evaluation also paid attention to what type of recommendations were more often preferred in a physical shopping scenario: either those where the products had similar technical specifications, or those where visual appearance was the base of the recommendations.

Throughout the different studies, the approach was generally well received in terms of usability, user experience, decision support and use intention, and was consistently perceived as useful and intuitive. As a whole, these studies offer insight into the suitability of AR for providing in-store product information (RQ2) and the implications of a hybrid shopping environment (RQ3). The specific contributions of each study are detailed in the following subsections, each dedicated to one of the aforementioned topics.

4.1. Acquisition of information via AR

The outcome of the first study indicates that including comparison tools in physical settings may allow for a quicker acquisition of information than when comparison features are not enabled. It was also revealed that, when comparing attributes, users prefer simple, absolute values rather than more elaborate approaches, such as the presentation of relative differences, which defeat their purpose by increasing the mental effort of evaluating them.

During the second study, participants made extensive use of the comparison and recommending functions, which may be an indicator of their perceived helpfulness. The results obtained in ResQue (Pu et al., 2011) (a questionnaire oriented to the evaluation of recommender systems) suggest that AR-based recommendations effectively support the discovery of new and diverse products. However, participants also called for better explanations of the recommendations.

Attribute exploration was also a focus of criticism. This issue seems to depend, to some extent, on current technology limitations and future advances, as new interaction methods and better AR displays can improve the navigational aspect of the approach.

4.2. Implications of a hybrid shopping setting

Based on the results of the second lab study, allowing clients to buy in a hybrid shopping environment (where products from a digital catalog can be compared to, and explored through, the physical ones) appears to have an effect on how users navigate the digital space and define their own preference models. Users seem to filter the product space intuitively: first, by focusing only on the physical products that match their preferences, and second, by exploring the items recommended in those products. In other words, their final decisions were mostly based on the preferred physical item for a certain task, which was used to further explore similar, digital possibilities *via* attribute critiquing. Furthermore, participants considered physical products as helpful for forming an opinion of the qualities of those presented only in digital form, also reflected in higher levels of purchase confidence, even if the selected products were not physically available.

5. Online user evaluations on user acceptance: Effects of psychological characteristics and comparison against a baseline system

Once the system had been developed and optimized in the user-centric process described above, the research focused on the general acceptance of AR-based in-store assistance. For this purpose, two main goals were established: first, to find the possible differences that may exist between types of consumers, and second, to obtain information on the acceptance of the developed approach in comparison with a baseline system (RQ4).

An online pre-study was performed to obtain an early assessment of the effects that psychological traits may have on user acceptance. The outcome of this exploratory pre-study suggested the existence of consumer types that differ in their perception of the system. To better discern whether these differences were caused by the use of AR features, or whether they were due to a more general disinclination toward the use of the technology, a further online evaluation was conducted comparing the AR-based prototype to a baseline system without AR.

Because the online pre-study is covered in more detail in a previous publication (Álvarez Márquez & Ziegler, 2021), only a summary of its results is presented in this article. The outcomes of the comparison study, however, are presented in full detail, as they are published here for the first time.

5.1. Online pre-study: Initial identification of differences among consumer types

In this pre-study, the same prototype developed during previous evaluations was used, with the goal of collecting data on user acceptance from a larger quantity of users. To that purpose, the concept and prototype for in-store AR-based shopping support systems were showcased through several videos that explained the main functionality. Together with questions concerning the acceptance of the system, the survey collected information about psychological traits of participants. These traits were measured through scales that assess relevant factors for the acceptance of in-store AR-based functions. Among the several possibilities contemplated to evaluate each factor, those that required participants to answer fewer items while still providing sufficient information to elaborate a reliable psychological profile were chosen, with the aim to maintain the overall length of the survey as short as possible. The considered psychological traits were:

- Technology acceptance: being open-minded toward technology and believing in its benefits may hold significance in the acceptance of the approach due to its high technological component, particularly regarding the use of head-mounted displays. Thus, the Technology Adoption Propensity index (Ratchford & Barnhart, 2012) was used to measure how probable it is for a person to adopt new technology.
- Shopping style: people who approach shopping differently may as well perceive the system in different manners, or give importance to different features, depending on which facets of shopping are more relevant to them. The Chronic Shopping Orientation scale (Büttner et al., 2014) was chosen to assess this trait, which helps distinguish task-focused shoppers from experiential ones.
- Decision style: supporting the purchase decision is the main objective of the concept here proposed. People have different ways of making decisions, and it may be that the system is tailored for some decision styles more than others. The Decision Styles scale (Hamilton et al., 2016) is used to obtain information on the extent to which a participant has a rational or intuitive decision style.

Data from 63 participants (40 females, average age of 34.1, σ 12.29) was collected and used to investigate whether different consumer types can be identified and the possible differences that may exist in their perception of the system, in order to gain insight into how different psychological profiles may influence the acceptance or rejection of in-store AR-based shopping support functions. All participants were recruited through the online platform *Prolific*[3] and received a monetary reward of £1.50 after successfully completing a survey (as per the site's policy). Demographic information is shown in Tables 1 and 2.

5.1.1. Results of the pre-study

Feedback received on the usefulness of AR-based shopping functions can generally be considered positive. The scores received for the different functions and interaction methods are all in the positive range. The constructs *perceived usefulness* and *decision support* received particularly high ratings. In agreement with previous evaluations of the system, participants found that the presence of physical products might be helpful for making a final purchase decision. However, among all the features of the system, direct inspection and interaction with real products was regarded as the less important one, surpassed by comparison support, product recommendation and access to a digital catalog.

Table 1. Demographic information of the sample.

Country	#
United Kingdom	43
United States	8
Ireland	4
Netherlands	2
Other (≤2)	6
Education level	**#**
Less than high school	2
High school	9
Bachelor's degree	19
Master's degree	32
Other	1

Table 2. Knowledge about augmented reality technology.

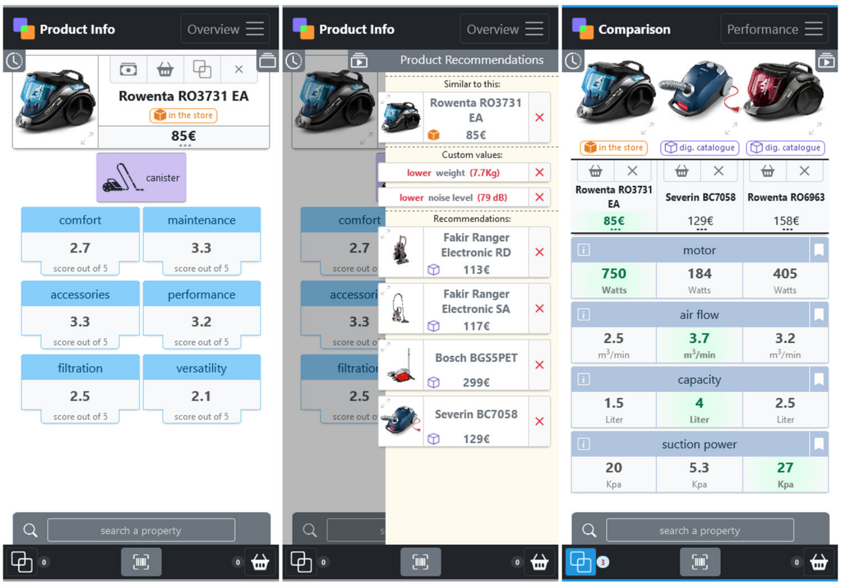
I know nothing about it	8
I know the name, but not much more	21
I know a bit about it and its possible applications	30
I have followed its development and know it well	4
I know a lot and could be considered an expert in the field	0

Despite the good overall results regarding system usability and user experience, the pre-study uncovered significant differences between consumer types in their assessment of the approach. Particularly, four well-distinguishable consumer types were found through clustering,[4] based on the psychological profiles of participants. While technology proficient users appear to be more likely to accept in-store services *via* AR like the ones described in this article, their level of exposure to AR seems to play a moderating role in that regard, that is, the more they know the technology, the less interested they are in using it. A possible explanation may be that most people are only exposed to AR with entertaining or advertising purposes, where its utilitarian value is low, which in turn may increase the user's disbelief in the real practicality of the technology. Higher acceptance was also observed in users who see shopping as an experience rather than a task. However, there was a general sentiment of concern about privacy and social acceptance factors, more pronounced in those participants with a strong intuitive decision-making style. These concerns have already been acknowledged in previous research as important obstacles for the acceptance of AR applications (Rauschnabel et al., 2018; Wassom, 2014).

5.2. Development of a non-AR-based smartphone app as baseline system

Despite the interesting results obtained in the pre-study, the lack of a baseline prevented the estimation of the extent to which the use of AR is related to discrepancies in acceptance among users with different psychological traits (RQ4). To better understand the role of AR technology and the implications of using it for providing in-store functions, it would be necessary to compare the prototype against a similar baseline system without AR capabilities. However, to date and to the best of our knowledge, no application exists that includes all the functions presented here through more conventional means (i.e., no AR involved) and, therefore, it was not possible to perform an adequate comparison prior to this study. In consequence, a smartphone app was developed to that purpose, keeping its functionality as close as possible to the original AR-prototype but without the AR features that it possesses. Despite the differences between head-mounted displays and smartphone devices, a smartphone was chosen as baseline because it was considered a realistic alternative technical option. Web-based technology is used to mirror the functionality of the AR-based prototype, but presenting the information similarly to a conventional online store and adapted to the small screens of smartphones. Nonetheless, a one-to-one conversion is hardly achievable due to how different an AR HMD approach and a non-AR smartphone one are in terms of information visualization and interaction possibilities. To better make use of the advantages of each technology and allow for a fairer comparison, some differences exist between AR- and non-AR-based implementations.

The new system allows the detection of physical products *via* QR scanning, after which the information relative to the scanned product is presented on the screen (Figure 6a). This information is organized exactly in the same way as it is done in its AR counterpart (categories, attributes and sub-attributes), and offers the same functions for explaining, bookmarking, and critiquing product characteristics, excluding showing any spatial-related information (e.g., the



a. Product information. "Overview" category is shown (scores per category).

b. Recommendations. Base product and some critiqued attributes have been defined.

c. Comparison view. Attributes in the "Performace" category are shown.

Figure 6. Different functions of the non-AR app.

location of a component in the product). Because product information is detached from the physical item, showing data relative to digital-only products (i.e., recommendations) is more easily achieved, and their attributes can be explored without distinction to those that belong to physically available items.

Recommendations are displayed four at a time, with the possibility to choose the product to base them on (Figure 6b). A difference in this regard is that to generate recommendations the base product must be explicitly chosen by the user, in contrast to the AR approach where recommendations are directly displayed and linked to the specific physical product within the user's field of view. On the other hand, this also provides a small advantage, because in the non-AR app it is possible to request recommendations based on other recommended products, which may be digital-only ones.

Regarding the comparison aspect, inspected products can be added to the comparison view, where up to three of them (digital or physical indistinctly) can be compared at once (Figure 6c). In that view, products and their attributes are displayed side-by-side, and best values are highlighted.

Altogether, when it comes to information visualization, each system has some inherent advantages in comparison to the other. While the AR-based one may have the benefit of a more immediate information acquisition and comprehension thanks to displaying superimposed, contextual data, the non-AR approach makes less distinction between digital and physical products, and permits a more structured and unified presentation of their attributes.

5.3. Online study: Comparison of the system against a baseline

To compare the AR system developed against a baseline using more conventional means of information provision, we conducted a new online user evaluation. Considering the exploratory nature of this comparative study, an online survey was chosen over a laboratory one because it allows for a quicker and larger collection of data, specially taking into account the still ongoing COVID-19 restrictions. Furthermore, the results in the pre-study (Álvarez Márquez & Ziegler, 2021) indicate that different types of consumers may also have contrasting opinions about the provision of AR-based functions in physical retailing, reason for which psychological data of participants was also collected this time. This data can be used to further investigate whether these discrepancies in the acceptance of the system exist due to the employment of AR HMDs, or it is the more general idea of using technology in a physical setting what makes the difference (RQ4).

5.3.1. Setting

The study followed a within subjects design with two controlled conditions: providing in-store functions either *via* an AR-HMD or a non-AR smartphone app. The first part of the survey included a series of questionnaires to collect psychological data of participants, which were to be used in the analysis of the results of this study and to extend/confirm the results obtained in the pre-study. Participants were then introduced to the concept of in-store shopping support systems through a video. After obtaining an overview of the approach, the implementation of different features (i.e., access to product information, comparison and recommendations) was described in more detail in three separate videos, which showed real prototypes for both AR and non-AR methods. At the end of each video showcasing a feature, participants were requested to score both systems in aspects related to what they just watched. At the final part of the survey, another questionnaire was presented that assessed more general factors.

5.3.2. Method

Sixty-four participants (36 females, average age of 30.53, σ 10.32) were recruited through the online survey platform *Prolific*,[5] and were rewarded £1.90 for their participation (as per the site's policy). Table 3 shows that around half of the participants were residents of the United Kingdom, and the majority had a Bachelor's degree or higher. In addition, they mostly had occupations that required user-level or advanced technology skills, but augmented reality was generally unknown to them (Table 4).

The survey was hosted in *SosciSurvey*.[6] The first part comprised the questionnaires:

- *Technology Adoption Propensity* (TAP) index (Ratchford & Barnhart, 2012), a 14-items-long questionnaire to assess how likely a person is to adopt new technology. It includes the sub-scales *optimism*, *proficiency*, *dependence* and *vulnerability*, which are rated using a 5-point Likert

Table 3. Demographic information of the sample.

Country	#
United Kingdom	27
South Africa	9
Portugal	7
Italy	5
Other ($\leq$3)	16
Education level	**#**
Less than high school	1
High school	17
Bachelor's degree	22
Master's degree	19
Other	5
Technological skills in current occupation	**#**
It requires advanced technological skills (e.g., programming).	16
It requires user-level technological skills (e.g., administrative software).	33
It requires little or no technological knowledge.	15

Table 4. Knowledge about augmented reality technology.

I know nothing about it	7
I know the name, but not much more	21
I know a bit about it and its possible applications	34
I have followed its development and know it well	2
I know a lot and could be considered an expert in the field	0

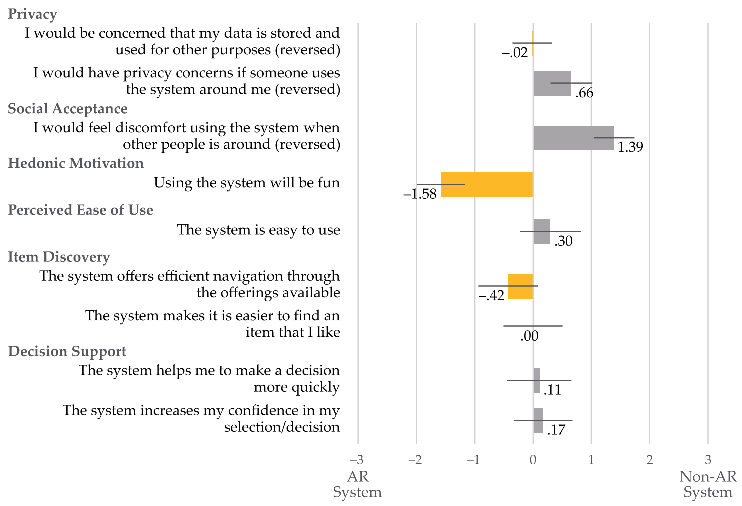


Figure 7. Average of preferred system per item. 95% confidence intervals. The original scale (1–7) has been changed for an easier interpretation of the results (–3 to 3). Negative values are in favor the AR-based system, while positive ones favor the non-AR baseline.

scale. The final score is the sum of the average scores of each factor (inhibiting ones are reverse coded).

- *Chronic Shopping Orientation* (CSO) scale (Büttner et al., 2014), to identify a consumer's disposition to be an experiential or a task-focused shopper. It uses a 7-points Likert scale that goes from *task-oriented* (lower values) to *experience-based* (higher values) shopping orientations.
- *Rational and Intuitive Decision Styles* (RDS and IDS) scale (Hamilton et al., 2016), to assess what type of decision-making style (rational or intuitive) is more prevalent. The outcome consists of a score for each style (in a 5-point Likert scale).

Right after filling out this information, participants were asked to watch a video explaining our general concept for in-store services, in which there was no explicit reference to the use of AR. Three more videos followed the introductory one, each focused on a specific functional aspect, that is, access to product information, recommendations, and comparison support. These clips showcased the functions with real prototypes for both AR and non-AR approaches, and were presented in counterbalanced order to control sequence effects. Each video also included a short questionnaire in relation to the specific function showcased. The final part of the survey comprised a short questionnaire to make a direct comparison between systems (this time taking into account their whole set of functions) by letting participants choose which one seemed more in line with a certain statement, considering the constructs: *decision support*, *item discovery*, *perceived ease of use*, *hedonic motivation*, *social acceptance*, and *privacy*. To that end, participants had to provide a value in a polar scale, with the AR-based approach at one extreme and the non-AR system at the other (total distance of seven points between them). Items about *intention to use* in two different scenarios were included too. The first scenario was that of a specialized store with products of interest for the participant; the second setting focused on special events, like trade-fairs or marketing actions. For each scenario, participants had to choose the system they believed was more fitting, along with their expectations about frequency of use.

The results obtained in each one of the aforementioned functional aspects and factors are presented in the next section. Besides the comparison between the AR system and the non-AR baseline, results concerning possible effects that particular psychological characteristics may have in their assessment are also reported.

5.3.3. Results

The results obtained for the questionnaires included after showcasing each functional aspect are listed in Appendix A. The means for each item were generally high for both systems and few significant differences were found. After a paired samples t-test and the adjustment of the obtained significance levels by means of the Benjamini-Hochberg procedure, it seems that the AR system was considered better suited for comparing products than the non-AR one, especially when it comes to the detection of differences between physical attributes (AR mean 4.33, σ 0.798; non-AR mean 3. 81, σ 1.052; $p < 0.05$) and understanding them. The recommendation aspect was perceived very similarly in both systems, as it was the access to product information. Only the effort of identifying products was different in that regard, for which the AR HMD was perceived as a more direct method for doing it (AR mean 4.30, σ 0.954; non-AR mean 3.84, σ 0.859; $p < 0.05$). The results also show that the smartphone was perceived to be slightly better in terms of preference elicitation, understanding of information, and navigational factors, although no statistically significant differences were observed.

The results of the polar questions are shown in Figure 7. They indicate that concerns about a subject's privacy when someone else uses the system nearby, as well as in relation to social acceptance when the subject itself uses it, are indeed present. The hedonic value of the AR system is

Table 5. Intention to use by setting.

	Store				Special event			
	AR		Non-AR		AR		Non-AR	
Never	0		1		2		1	
Occasionally	10		24		25		12	
Frequently	10		19		15		9	
Total	20		44		42		22	
Functions (1–5 scale)	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Non-AR smartphone								
Product information	4.00	0.163	4.32	0.077	4.30	0.089	4.07	0.133
Product comparison	3.75	0.172	4.03	0.089	3.92	0.092	3.99	0.167
Product recommendations	3.89	0.167	4.17	0.091	4.14	0.087	3.97	0.173
AR headset								
Product information	*4.48	0.114	*4.05	0.095	⋆4.32	0.089	⋆3.92	0.137
Product comparison	*4.43	0.109	*3.99	0.101	⋆4.26	0.102	⋆3.88	0.118
Product recommendations	**4.50	0.117	**3.93	0.103	4.21	0.103	3.91	0.148
Constructs[a]	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Decision support	**1.45	0.357	**0.86	0.231	0.14	0.287	0.68	0.393
Item discovery	**1.57	0.269	**0.41	0.253	0.51	0.271	0.36	0.375
Perceived ease of use	**1.50	0.373	**1.11	0.242	⋆0.07	0.307	⋆1.00	0.411
Hedonic motivation	1.90	0.347	−1.43	0.242	⋆2.00	0.196	⋆−0.78	0.394
Social acceptance (rev.)	−1.05	0.246	1.545	0.214	−1.19	0.187	1.77	0.322
Privacy (rev.)	−0.15	0.262	0.398	0.184	−0.25	0.165	0.45	0.307

Following the data on use frequency, scores per groups based on system choice (AR or non-AR) in each scenario (store or event) are reported, firstly concerning the functions of each system (scored using a 1–5 scale), and secondly the constructs addressed in the polar scale (while original values from 1 to 7 were used in the calculations, the results have been re-coded to support an easier interpretation—see table footnotes).

[a]Original polar scale 1–7, 1 = AR Headset, 7 = Non-AR Smartphone. Re-coded to 3 to −3 for partici pants who chose AR, and −3 to 3 for those who preferred non-AR.

*$p < 0.05$, **$p < 0.01$ (Benjamini-Hochberg correction)//⋆$p < 0.05$ (uncorrected).

clearly higher than that of the non-AR approach, and it also seems to be perceived as a better way to explore the products available at the store. Nonetheless, both systems are seen as equivalent in most other aspects, that is, personal data protection, ease of finding a proper product, and decision support. Furthermore, the non-AR system is believed to be only marginally easier to use than the AR one.

Data on intention to use can be found in Table 5. Based on it, AR HMDs are generally perceived as more suitable for event-like contexts, such as trade fairs or marketing actions, while the smartphone seems more adequate for day-to-day shopping. The same table reports data on scores given to each functional aspect and construct by chosen system and setting (store or special event). Independent samples Mann–Whitney *U* tests per setting type show that in both scenarios the perception that participants have of the functions provided by the AR HMD is a more decisive factor than that of the smartphone. Within the store scenario, people who chose AR over the non-AR approach gave higher values than those who didn't in terms of how they perceived the *product information* ($p < 0.05$), *comparison* ($p < 0.05$), and *product recommending* ($p < 0.01$) capabilities of the AR HMD. Differences between groups also appeared to some extent in the special event setting for *product information* and *comparison* with the AR prototype, but did not pass the significance levels correction procedure. *Decision support*, *item discovery*, and *perceived ease of use* also seem relevant when deciding which system to choose in a store (all $p < 0.01$). The results also suggest the possibility of *perceived ease of use* ($p < 0.05$, uncorrected) and *hedonic motivation* ($p < 0.01$, uncorrected) to drive the selection of one system over the other in an special event scenario. The perception of the smartphone app has no impact on which system is chosen, no matter the setting, and the same occurs with social and privacy constructs. Also in this matter, a final Fisher's exact test[7] was performed to examine the relation between chosen system and expected frequency of use, which resulted in a non-significant outcome; that is, there is no relation between what system is chosen and how often it would be used in any of the scenarios.

5.3.4. Influence of psychological characteristics

To analyze the influence of psychological characteristics on the perception and acceptance of the AR system, the data obtained in this study were pooled with a sample collected in a previous one (Álvarez Márquez & Ziegler, 2021). The analysis mirrored the two-step process used in that previous study: first, a hierarchical cluster analysis was conducted to classify the data based on average linkage between groups, whose outcome provided an initial distribution of participants and outliers; second, a K-means clustering analysis was performed to corroborate those first results. Z-scores of the variables were calculated and used in the analysis due to their different original scales. The four resulting clusters had very similar psychological traits as the ones determined by the previous, smaller sample. Participant distribution of both samples merged can be seen in Appendix C (127 subjects).

Concerning only the sample of this study, five participants were considered outliers, leaving a total of 59. Trait scores of these remaining subjects per resulting groups can be seen in Table 6. As a summary of their predominant characteristics, group 1 is formed by participants who are more intuitive and experiential shoppers than the average of the sample; group 2 contains technology adopters; group 3, on the other hand, has subjects who are less likely to adopt

Table 6. Psychological traits per group.

	#	Fem.	Age	AR Kno.	TAP Sub-scales Opt.	Pro.	Dep.	Vul.	TAP Score	CSO Score	RDS Score	IDS Score	RDS - IDS
G1	21	67%	28.29	0.18	0.04	-0.12	0.35	0.33	-0.84	1.03	0.15	0.35	1.06
G2	11	27%	30.27	0.24	0.38	0.49	-0.75	-0.69	2.31	-0.67	-0.28	-0.15	1.13
G3	12	50%	37.17	-0.32	-0.40	-0.48	0.22	0.10	-1.00	-0.91	-0.56	0.03	0.67
G4	15	60%	28.00	-0.02	0.08	0.06	-0.24	0.16	0.22	-0.77	0.50	-0.35	2.11
Total	59		Avg.	2.48	4.38	3.94	2.78	3.57	13.97	3.757	4.22	2.97	1.25

From left to right: group size; percentage of females; average age; AR knowledge (on a 1–5 scale); TAP sub-scales (1–5): Optimism, Proficiency, Dependence, Vulnerability; total TAP score; Chronic Shopping Orientation (1–7); Rational and Intuitive Decision Styles (1–5); and the difference between them. Values of the psychological traits are relative to the average of the sample without outliers (last row of the table). Colored cells identify those values where a group's mean noticeably differs from the total.

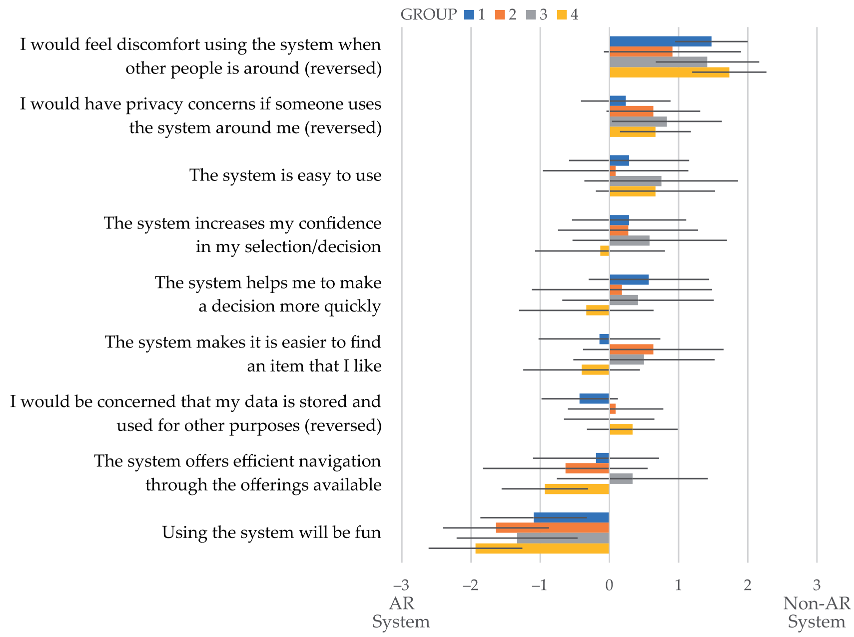


Figure 8. Average of preferred system per item and group. 95% confidence intervals. The bars to the left are in favor of the AR-based system, while the bars to the right are in favor of the non-AR one.

new technologies and present a lower rational decision style, as well as the most task-focused shopping orientation; finally, group 4 is the most rational and least intuitive one. Besides, group 2 is the more knowledgeable about AR, while subjects in group 3 know the least.

The full set of scores per group and functional aspects are reported in Appendix B. There is an indication that different consumer types have different perceptions of the approach. In general, groups 1 (intuitive, experiential shoppers) and 4 (most rational) have a more positive view of the presented in-store functions, while group 3 (the lowest technology proficiency, most task-focused shoppers) shows lower scores than the rest. Groups 3 and 4 show a difference in their perception of how conveniently product attributes are presented in the AR HMD ($p < 0.05$) and how useful the smartphone is for learning a product's functionality ($p < 0.05$), in both cases being group 4 the more positive one. Group 3 also presents differences with groups 1 and 4 when it comes to the comparison of physical attributes in the AR system ($p < 0.01$ and $p < 0.05$, respectively, for which groups 1 and 4 are more positive than 3), and with group 1 only for the comparison of digital attributes in the non-AR approach ($p < 0.05$, group 1 with higher values). Moreover, there is a difference between groups 1 and 3 for the overall comparison aspect in the AR HMD ($p < 0.05$), for which group 1 gave better scores.

Despite the variation in scores observed between groups in their evaluation of the two systems, after performing the corresponding paired samples *t*-tests and using the Benjamini-Hochberg correction, no statistically significant differences were found in how a group individually assessed the AR HMD in comparison to the non-AR smartphone.

Figure 8 shows the results of the polar questions per group. There is an indication that groups 1 and 4 are more positive about the benefits of AR HMDs, while group 3 will prefer the non-AR smartphone version most of the time.

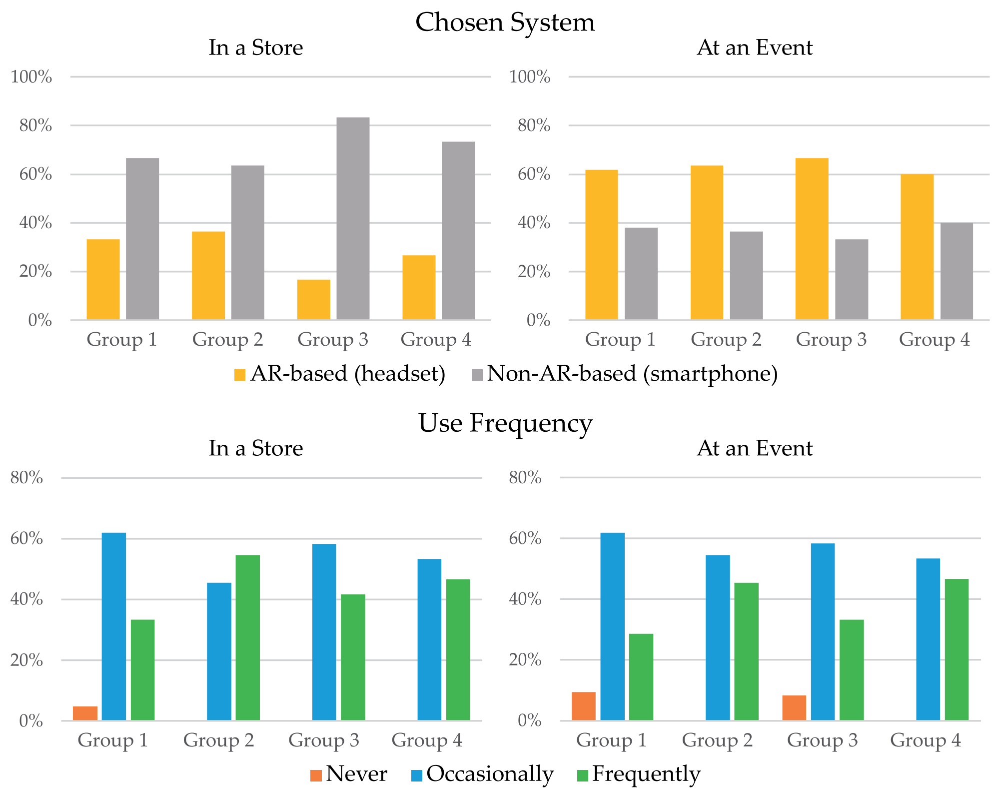


Figure 9. Chosen system and expected use frequency of in-store functions (through either system) per group and scenario.

Nonetheless, the scores appear to be mostly aligned between the groups and no significant differences were found.

Finally, as to the intended use of each system in each of the proposed scenarios (a store or a special event), a Fisher's exact test reported no significant association between groups and system of preference. Similarly, no association was found between belonging to a certain group and the expected frequency of using in-store services in either setting. A comparative chart of chosen system and use frequency, per group and setting, can be seen in Figure 9.

5.3.5. Discussion

When considering all participants as a whole, the results show little difference between both implementations of our approach to in-store shopping support. As expected, some of these differences are inherent to their platforms, such as the perception of a faster acquisition of information in an AR HMD thanks to digital superimposition, or the anticipation of an easier navigation offered by the already well-known interaction on smartphones. Still, the value that AR holds in physical retailing (RQ2) is highlighted one more time, not only in how participants think of it as the more entertaining solution, but also as the one that provides a better view of the product space and physical comparison. That the systems are perceived very similarly in other aspects is not necessarily a negative outcome, considering that smartphones are already a part of everybody's life and can be used as a good reference to measure the acceptance of a new technology. In other words, within the studied context, providing in-store functions *via* AR through HMD seems to provide some clear advantages in comparison to a non-AR based approach for smartphones, while in most other aspects the systems appear to be on par, except for some issues raised in relation to privacy and social acceptance. Nonetheless, these two factors seem to have no influence in how users decide which system to use (given the opportunity), regardless of the setting: on the one hand, participants appear to base that decision on their perception of the functions of the AR HMD alone, no matter what they think about the smartphone approach; on the other hand, in a store scenario, the perceived pragmatic qualities of the systems seem to be the driving force behind deciding for one or the other, while in a special event situation, the hedonic qualities have more weight instead.

Regarding types of consumers, the results are in line with previous evaluations, where the psychological traits seem to influence to some extent the perception that users have of the systems (RQ4). Interestingly, consumers with apparently opposing characteristics are closer in their evaluation of the in-store functions presented here. There is an indication that both rational, task-focused shoppers and intuitive, experiential ones, are more positive about the benefits of using in-store services, either *via* an AR HMD or a non-AR smartphone. On the other hand, consumers with low technology proficiency and no previous knowledge about AR are more negative toward either system, while people with high technology acceptance propensity and knowledgeable about AR also seem to be less enthusiastic regarding the AR headset, and even favor the non-AR smartphone approach above it. This outcome also agrees with previous studies, and a possible explanation may be that these users, despite knowing AR and having tried it in the past, were rarely exposed to applications with utilitarian purposes and, thus, they might not see its practical value. Nonetheless, no significant differences were found in how each group, individually, scored the systems. Participants with high intuitive decision-making and experiential shopping orientation go one step further by also consistently evaluating both systems similarly. The lack of significant differences may indicate that the rejection of the AR system could be related to a broader factor (such as a general aversion to technology utilized for the purpose here described), rather than to the more specific use of AR HMDs (RQ4). The effects of the different psychological traits do not seem to have a significant impact on the use intention of either system, and groups appear to agree on the suitability of AR for special events, while generally preferring the smartphone in a physical store situation.

5.4. Limitations of the study

Participants did not test the systems themselves, but only experienced them through a series of explanatory videos. That means that the obtained results may only be valid to better understand why a new user may choose one system over the other, but are not enough to assess the acceptance of in-store AR functions in the long run and after having used both systems. More specifically, users may score very differently those aspects of the approach that heavily rely on experience and for which users may not have previous references (such as navigation in an AR HMD) after actually getting to use the prototypes.

6. Interactive evaluation: User experience and usability against a baseline

While initial studies of this research were more of an informative nature, and later ones focused on comparing the prototype against a baseline system from a conceptual point of view (participants could not interact with the systems), there was still a lack of real usage data with which to assess the performance of the developed AR approach in comparison to a more established information retrieval method. Consequently, a complementary, smaller laboratory experiment was conducted to fill this gap, focused on obtaining more information about differences in the usability and user experience of the systems.

6.1. Setting

The study took place in a room with three shelves, each one containing a physical vacuum cleaner. The selected models covered different usage areas, but were sufficiently similar so that comparing them made sense (a. bagless, battery-powered stick vacuum cleaner, b. bagless, small, standard canister vacuum cleaner, and c. bagged, big, wet-dry canister vacuum cleaner). The source of the digital recommendations was a database with 100 vacuum cleaners, including the three physical ones.

The study followed a within-subjects counterbalanced design, where the manipulated variables were the use of AR HMD and non-AR smartphone approaches. Although physical differences exist between devices, the same functions with same available information were implemented in both cases, and the study focused on the differences between the acquisition of such information by one or the other method. Participants had to test both systems by trying to find an adequate product for themselves or a family member. After testing a system, they were asked to fill questionnaires concerning *specific functional aspects*, *usability* and *user experience*. By the end of the session, participants answered some more open questions, where they were able to provide further insight into their thoughts about both systems.

6.2. Method

Thirteen participants took part in the lab experiment (seven female, average age of 27.54, σ 5.08), all of them German residents and mostly students with a technological background. At the beginning of a session (which was individual for each subject) participants were informed about the scope of the research, and were asked to formally consent to the use of the data collected. Afterwards, they were requested to explain AR with their own words. Only five declared to have tried AR before, and up to seven were able to correctly describe it; two participants, however, thought that AR and VR were the same, and the rest (4) could not provide a proper description.

After a brief discussion concerning their view of future retailing, participants were explained how to use one of the prototypes and were given some time to get used to it, whereupon they had to use it to find a suitable product for themselves. Upon completion of the task, they filled three questionnaires:

- One with the same items used in the previous online study to assess functional aspects related to *information acquisition*, *product comparison*, and *product recommending*; this last aspect was extended with the addition of two new questions about the user's final product selection.
- The *System Usability Scale* (SUS) (Brooke, 1996), which provides a score by which the usability of different systems can be compared. It consists of 10 items and uses a 5-point Likert scale.
- The short version of the *User Experience Questionnaire* (UEQ-S) (Schrepp et al., 2017), an 8-items long questionnaire to evaluate hedonic and pragmatic qualities of a system. Each item presents a pair of terms with opposite meanings at each end of a 7-point Likert scale (−3 to 3).

Next to testing the first prototype, the same procedure was followed with the other one, but requiring to find a product for a family member instead (counterbalance

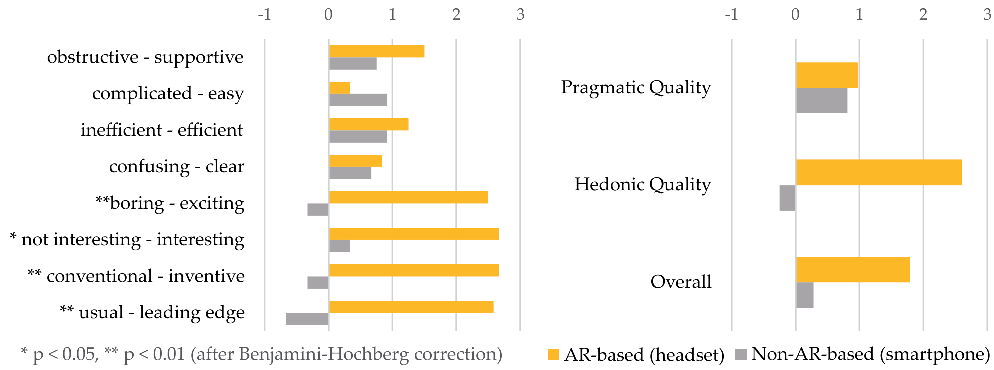


Figure 10. UEQ-S results for each item and construct. The scales have been shortened for readability.

measures were taken to decide the order in which the systems were tested). At the end, participants had the opportunity to freely express their feelings about each system and their experience, as well as to explain which one was preferred.

6.3. Results of the questionnaires

Both systems scored very similarly in terms of *usability*, but the smartphone version was slightly better regarded. The average SUS score for the AR HMD was of 59.81 (σ 19.89), and the non-AR smartphone obtained 64.62 (σ 19.34), which would qualify both of them as "OK." A paired samples *t*-test corrected by the Benjamini-Hochberg procedure reported no differences between the means of the items of the SUS.

Results of the UEQ-S are reported in Figure 10. The systems were mostly considered to be at the same level regarding their pragmatic quality, only showing a more notable difference in terms of complexity (for which the non-AR smartphone obtained better scores) and support (in favor of the AR HMD). However, when it comes to hedonic quality, the HMD was clearly perceived as the better option and outperformed the smartphone in all the items of this factor.

The scores that each system obtained concerning their specific functional aspects are presented in Appendix A, next to the ones collected during the previous online study. The scores of both systems tended to be lower in the lab study than in the online one (were the systems could not be tested), although this trend seems to be more accentuated in the case of the non-AR smartphone. Despite this, it is possible to see an improvement between studies in the scores of the AR approach in most items related to the access of product information. Some differences between the systems were also made more apparent in the lab experiment, as it was the case for the *ease of exploring product attributes*, *how conveniently these attributes are explored* ($p < 0.05$), and *how easy it is to find/see/identify product recommendations* ($p < 0.05$), all for which the AR HMD was the better regarded approach. Altogether, it appears that after testing both systems participants found the AR headset to be better for retrieving and exploring product information, learning about it, detecting differences between products (not matter whether these are physical or not) and accessing recommendations; on the other hand, the smartphone seems to be better considered for knowing which product is closer to the preferences of the user.

6.4. Answers to open questions

About the future of retailing, and previous to trying any system, there was a generalized opinion among participants that online-based transactions would take over traditional commerce, meaning that online shops, online communication with sales personnel, and technologies that enable trying products from home (like VR) will be the norm. As for physical retailing, frequent comments were related to how technology will replace humans (e.g., cashiers) and that stores will be more like places to try things, not to purchase them (showrooms).

When comparing the systems, general thoughts were that the non-AR smartphone version was easier to use and more convenient (mainly because they owned one and already knew how to use it), and that comparing attributes seemed faster and easier because of the well-known table view. However, there was a downside to this, that is, many stated that they had higher standards for the smartphone app than for the HMD one. Other common concerns were related to the detachment of information from physical products, which forces users to stay focused on the screen of the phone. Some argued that this detracts from the experience of shopping in a physical store, as it removes the main purpose of going to one, that is, interacting with products and other people.

The AR HMD approach was typically described as novel, fun and/or exciting, and most participants expressed surprise about the interaction possibilities (i.e., AR UIs with interactable elements similar to those in more standard displays). They often highlighted the comparison of physical attributes and the readiness of the information as its main benefits. The approach was found to be more suitable for physical retailing, as it does not lose focus from physical products and encourages their exploration. Nonetheless, some participants declared to be overwhelmed by the quantity of the information, and that they could not imagine how it would look like in a real setting with a larger set of physical products. A majority of participants stated that it was required to learn a lot to use the system properly, highlighting issues like that they would need help to use it or that navigating through the different options was complicated (e.g., head-gaze pointing or locating digital objects). However, most participants claimed to be willing to invest the time to learn it given the opportunity. Privacy was not considered an issue, under the argument that smartphones already collect privacy-sensitive information (e.g., fingerprints); but for many wearing an AR HMD would not be socially acceptable, as long as they are such a conspicuous device or their use is not more spread.

Lastly, to the question about which one they would use right now, four participants chose the AR headset mainly because of its novelty and entertainment factors, and what they thought it was a better representation of product information. Another four preferred the non-AR smartphone on the grounds of a greater familiarity with the device and ease of understanding the information. The remaining five would use one or the other mostly depending on two factors: time and quantity of products. The time was considered a constraint for getting used to the AR system, as well as to explore products in detail. As for the quantity of products, when there are just a few or they have already been filtered beforehand, they argued that it is not worth the hassle of wearing the glasses and the smartphone is a more direct method. However, when presented with a large set of products, or when they know little about them, and thus require more exploration, the AR approach is preferred.

6.5. Discussion

In addressing our first research question, our developments demonstrate how AR can be leveraged in a hybrid physical-online shopping context by providing recommendation and comparison functions. We have iteratively refined the design of such system in a user-centered fashion obtaining relevant insights concerning which information to provide through AR and how to present it.

In answering our second and third research questions (greatly related in this case, as the advantages of using AR in a physical setting involve learning and exploring in a hybrid a space), online and lab studies suggest that using AR HMDs to provide in-store functions may have greater value in terms of product exploration and discovery than a non-AR smartphone approach, without showing a significant loss concerning other utilitarian factors, but a clear advantage in hedonic ones. A more dynamic and direct access to information seem to make AR HMDs the better platform to support learning about the product space, especially when a large number of choices are available. Moreover, providing in-store services through AR HMD also seems to be more appropriate for brick-and-mortar stores thanks to not losing focus from the physical space that surrounds the user, which in turn encourages product inspection, together with what appears to be a better presentation of differences between physical products. However, the implemented system for AR HMD seems limited when it comes to attribute presentation, as it appears to offer a less structured and clear view that the non-AR smartphone approach. This could be the reason behind the preference of the latter for dealing with small sets of products where exploration is not a relevant factor, and for which the assessment of their qualities and how close these are to the user's own preferences appear to be more easily achieved with a traditional table view.

In relation to the usage of dedicated AR-hardware (fourth research question), and in concordance with the results of the online study, social acceptance is considered one of the main challenges for adopting the use of AR HMDs in physical shopping environments, mostly due to the aspect of the device and how the user may look like when wearing it. However, the learning curve of using AR HMDs appears to be an even higher limitation. For the majority of people, AR HMDs are still a big unknown, and few are those who have ever been exposed to them, or that are fully aware of their possibilities. This is even more evident after recognizing that most of the participants in the lab study had a technology background, yet few were actually well versed in the current state of AR technology. This implies that AR HMDs are probably something new to most users, which may make them interesting or fun to try, but also intimidating and hard to use without proper guiding.

6.6. Limitations of the study

AR HMDs and smartphones greatly differ in their characteristics (e.g., comfort, interaction methods, or response times). While there was no appreciable performance differences between devices due to the limited number of products and the fact that most of the computational effort was performed on the server side of the application (which was the same in both approaches), it is true that head-mounted displays were generally perceived as less comfortable to use during long periods due to their weight and operability (e.g., head-gaze pointing). Although participants of the lab experiment were asked to focus on the concept and disregard these limitations, as they were outside the scope of the investigation, and the functionality and information provided by both systems were controlled to be as close as possible, these aspects may still have influenced their responses. Furthermore, different design choices could have been made for the implementation of the non-AR smartphone. Although both systems are very close feature-wise, and efforts have been made to perform a comparison as fair as possible, the higher standards that participants held for the smartphone app may have had an impact on the results. This can be noticed in the difference in scores observed between online and lab studies, which could be taken as an analogy for expectation *vs.* reality. Furthermore, only a reduced number of participants could take part in the study due to COVID-19 restrictions. Given the relatively large confidence intervals and the small sample, it is not possible to make strong statements about the collected data. However, answers given by participants to the more open questions seem to agree with what was inferred from the questionnaires and support the interpretation that has been made of them.

6.7. Implications for designers of in-store shopping experiences

AR is an increasingly relevant technology, and is currently receiving considerable support from leading companies. As a result, AR headsets can be expected to become a more accessible device, opening up exciting opportunities as enablers of omni-channel experiences. Withing an in-store context, designers can benefit of the multiple advantages that the technology offers in terms of information accessibility, product discovery, and customer engagement. However, AR is still a big unknown for most people, thus the significance of creating complex-free, enjoyable experiences to create a growing base of in-store AR consumers. To that end, it appears to be good practice to avoid showing floating, unconnected information, but to anchor it to real world objects that help users finding it when needed. Also important is to keep all relevant data contained within the field of view of the user, which supports its faster and easier processing, more so when dealing with data that is usually inspected together (like product properties that need to be compared for their assessment). To fully take advantage of AR, the surrounding environment should act as an interactive complement of the digital one, where enhanced exploration and navigation can work both ways: from providing digital cues to navigate the real world, to using real objects to intuitively explore the digital space. Finally, information must have a purpose, appearing only when necessary and not cluttering the real world so that this is obscured and relegated to a second plane, because, in a physical-store setting, touching and experiencing products may be as equally important for making a purchase decision.

6.8. Implications for retailers

Consumers expect more from future physical retailing than simply being a place where transactions happen, and predict more in-store digital services, personalized attention and less hassle in general. Retailers should start taking actions in that regard sooner than later, and begin designing new methods to attract clients and create unique experiences that make the trip to the store worth the effort. AR has plenty to offer to both retailers and consumers, and it would be wise keeping an eye on how it develops in future years. Most consumer types seem to have a general interest in, at the very least, trying and knowing more about this type of technology, which appears not only to be noteworthy as entertainment provider, but also has the power to bring pragmatic benefits that may rival those of more traditional technologies.

As of today, AR HMDs do not appear to be sufficiently developed for wide public adoption, while a large-scale deployment by retailers would still require a high financial investment in terms of hardware acquisition, setup, and staff training. Nonetheless, pioneering companies can start experimenting on a smaller scale and create store concepts that include novel on-site digital functions, where the line between online and physical channels blurs even further. That would mean providing clients with technology-based, tailored on-site information and services, while retailers may increase their profits by using their resources more efficiently. It is foreseeable that social and privacy concerns may become less acute in a close future, when AR HMDs evolve into less evident devices and more people start wearing them. For the time being, companies could gradually incorporate AR into their business model by designing private spaces, focused on specific products and advertising events, where clients can be introduced to AR in a more intimate manner. Actions like these may be of significance for understanding consumer needs and expectations, which could be employed to create more intuitive and easy to use AR UI's, as well as to gather valuable experience and gain advantage over less adventurous competitors. This type of events may as well increase people's awareness of AR technology and thus, its acceptance.

7. Conclusions

The use of augmented reality technology for the provision of in-store shopping services is a promising yet still underexplored research area. Gaps exist regarding the extent to which the use of AR, paired with recommender systems, supports the gathering of information and purchase decision of high complexity products when both digital and physical alternatives are available, as well as concerning the general acceptance of these type of in-store functions and how it is affected by customers' psychological traits. Here we propose a concept for AR-based in-store shopping assistants that provides extended product information, recommendations and comparison support. The inclusion of recommendations of products not physically available at the store creates a novel environment where physical and digital products coexist, with the implications that it may pose on how consumers explore and learn about them. The concept was implemented into a prototype for head-mounted displays (Microsoft HoloLens), which was used in a series of evaluation studies to assess the impact of its singular characteristics on the shopping experience, and address possible limiting factors for its acceptance.

The outcome of the studies suggest that the proposed in-store functions can create beneficial new dynamics in how consumers learn about, explore, and discover products. These studies also provide evidence of the existence of stable consumer types with different psychological traits, and the acceptance of in-store assistants in general seems to differ between them. The results also indicate that providing these on-site services *via* AR HMDs maintains and even improves the pragmatic qualities of using more established platforms, while clearly outperforming them in the hedonic ones. Finding utilitarian purposes for AR seem to be a key element for its acceptance, since only the practical aspects of the approach play a role in the user's intention to use the system at stores. This is further highlighted in the importance given to the ease of use and usability aspects of the system when deciding whether to make use of it or not, and it is hinted that these factors may pose a greater challenge in the adoption of AR HMDs than the more frequently addressed ones of privacy and social acceptance. Because the issues related to these two last factors are likely to be progressively mitigated the more the technology advances, it seems more urgent for current research to focus on defining efficient, easily understandable information visualization and interaction methods for HMDs, as these appear to be their biggest limitations in comparison to standard displays. In the time being, companies should focus on the creation of experimental environments: private spaces where consumers can learn about and become familiar with AR, while retailers may gather information about their needs and explore new shopping concepts in order to be better prepared for a future that is closer than ever.

Notes

1. www.ptc.com/en/products/vuforia
2. deepai.org
3. www.prolific.co
4. These results were expanded with data collected in the next study, using a total of 127 subjects. The new clustering results (reported in Appendix C) support the conclusions obtained in the pre-study.
5. www.prolific.co
6. www.soscisurvey.de
7. This test was chosen because it is more appropriate when the expected cell counts are small.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially supported by European Regional Development Fund, under Grant [EFRE-0801448].

References

Ahn, J., Williamson, J., Gartrell, M., Han, R., Lv, Q., & Mishra, S. (2015). Supporting healthy grocery shopping via mobile augmented reality. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *12*(1s), 1–24. https://doi.org/10.1145/2808207

Aisle 411 & Tango (2014). Retrieved September 3, 2022, from http://aisle411.com/

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Álvarez Márquez, J. O., & Ziegler, J. (2019). Augmented-reality-enhanced product comparison in physical retailing. In *Proceedings of Mensch Und Computer 2019, MuC'19* (pp. 55–65). Association for Computing Machinery. https://doi.org/10.1145/3340764.3340800

Álvarez Márquez, J. O., & Ziegler, J. (2020). In-store augmented reality-enabled product comparison and recommendation. In *Fourteenth ACM Conference on Recommender Systems, RecSys '20* (pp. 180–189). Association for Computing Machinery. https://doi.org/10.1145/3383313.3412266

Álvarez Márquez, J. O., & Ziegler, J. (2021). Acceptance of an AR-based in-store shopping advisor – The impact of psychological user characteristics. In *Human-Computer Interaction – INTERACT 2021, Lecture Notes in Computer Science* (pp. 457–479). Springer International Publishing.

Beck, M., & Crié, D. (2018). I virtually try it … I want it! virtual fitting room: A tool to increase on-line and off-line exploratory behavior, patronage and purchase intentions. *Journal of Retailing and Consumer Services*, *40*, 279–286. https://doi.org/10.1016/j.jretconser.2016.08.006

Beck, N., & Rygl, D. (2015). Categorization of multiple channel retailing in multi-, cross-, and omni-channel retailing for retailers and retailing. *Journal of Retailing and Consumer Services*, *27*, 170–178. https://doi.org/10.1016/j.jretconser.2015.08.001

Bollen, D., Knijnenburg, B. P., Willemsen, M. C., & Graus, M. (2010). Understanding choice overload in recommender systems. In *Proceedings of the Fourth ACM Conference on Recommender systems – RecSys '10*. ACM Press. https://doi.org/10.1145/1864708.1864724

Bonetti, F., Pantano, E., Warnaby, G., Quinn, L., & Perry, P. (2019). Augmented reality in real stores: Empirical evidence from consumers' interaction with AR in a retail format. In *Augmented reality and virtual reality* (pp. 3–16). Springer.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Büttner, O. B., Florack, A., & Göritz, A. S. (2014). Shopping orientation as a stable consumer disposition and its influence on consumers' evaluations of retailer communication. *European Journal of Marketing*, *48*(5/6), 1026–1045. https://doi.org/10.1108/EJM-04-2012-0210

Buytaert, D. (2018). *Shopping with augmented reality*. Retrieved September 3, 2022, from https://dri.es/shopping-with-augmented-reality

Caboni, F., & Hagberg, J. (2019). Augmented reality in retailing: A review of features, applications and value. *International Journal of Retail & Distribution Management*, *47*(11), 1125–1140. https://doi.org/10.1108/IJRDM-12-2018-0263

Chatzopoulos, D., Bermejo, C., Huang, Z., & Hui, P. (2017). Mobile augmented reality survey: From where we are to where we go. *IEEE Access*, *5*(99), 1. https://doi.org/10.1109/ACCESS.2017.2698164

Chen, C.-C., Huang, T.-C., Park, J. J., & Yen, N. Y. (2015). Real-time smartphone sensing and recommendations towards context-awareness shopping. *Multimedia Systems*, *21*(1), 61–72. https://doi.org/10.1007/s00530-013-0348-7

Chylinski, M., De Ruyter, K., Sinha, A., & Northey, G. (2014). *Augmented retail reality: Situated cognition for healthy food choices*. ANZMAC (Australian and New Zealand Marketing Academy).

Cruz, E., Orts-Escolano, S., Gomez-Donoso, F., Rizo, C., Rangel, J. C., Mora, H., & Cazorla, M. (2019). An augmented reality application for improving shopping experience in large retail stores. *Virtual Reality*, *23*(3), 281–291. https://doi.org/10.1007/s10055-018-0338-3

Dacko, S. G. (2017). Enabling smart retail settings via mobile augmented reality shopping apps. *Technological Forecasting and Social Change*, *124*, 243–256. https://doi.org/10.1016/j.techfore.2016.09.032

de Souza Cardoso, L. F., Mariano, F. C. M. Q., & Zorzal, E. R. (2020). A survey of industrial augmented reality. *Computers & Industrial Engineering*, *139*, 106159. https://doi.org/10.1016/j.cie.2019.106159

Dillon, A. (2001). *User acceptance of information technology*. Taylor and Francis.

Doherty, N. F., & Ellis-Chadwick, F. (2010). Internet retailing: The past, the present and the future. *International Journal of Retail & Distribution Management*, *38*(11/12), 943–965. https://doi.org/10.1108/09590551011086000

Enders, A., & Jelassi, T. (2000). The converging business models of internet and bricks-and-mortar retailers. *European Management Journal*, *18*(5), 542–550. https://doi.org/10.1016/S0263-2373(00)00043-8

Fagerstrøm, A., Eriksson, N., & Sigurdsson, V. (2020). Investigating the impact of Internet of Things services from a smartphone app on grocery shopping. *Journal of Retailing and Consumer Services*, *52*, 101927. https://doi.org/10.1016/j.jretconser.2019.101927

Gao, F., & Su, X. (2019). New functions of physical stores in the age of omnichannel retailing. In S. Gallino & A. Moreno (Eds.), *Operations in an omnichannel world, Springer series in supply chain management* (pp. 35–50). Springer International Publishing.

Gentner, D., & Medina, J. (1997). Comparison and the development of cognition and language. *Cognitive Studies*, *4*(1), 1_112–1_149. https://doi.org/10.11225/jcss.4.1_112

Georgel, P., Schroeder, P., Benhimane, S., Hinterstoisser, S., Appel, M., & Navab, N. (2007). An industrial augmented reality solution for discrepancy check. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR '07* (pp. 1–4). IEEE Computer Society.

Grohmann, B., Spangenberg, E. R., & Sprott, D. E. (2007). The influence of tactile input on the evaluation of retail product offerings. *Journal of Retailing*, *83*(2), 237–245. https://doi.org/10.1016/j.jretai.2006.09.001

Gutiérrez, F., Htun, N. N., Charleer, S., De Croon, R., & Verbert, K. (2019). Designing augmented reality applications for personal health decision-making. Accepted: 2019-01-02T23:56:24Z.

Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment*, *98*(5), 523–535. https://doi.org/10.1080/00223891.2015.1132426

Hilken, T., de Ruyter, K., Chylinski, M., Mahr, D., & Keeling, D. I. (2017). Augmenting the eye of the beholder: Exploring the strategic potential of augmented reality to enhance online service experiences. *Journal of the Academy of Marketing Science*, *45*(6), 884–905. https://doi.org/10.1007/s11747-017-0541-x

Hilken, T., Heller, J., Chylinski, M., Keeling, D. I., Mahr, D., & de Ruyter, K. (2018). Making omnichannel an augmented reality: The current and future state of the art. *Journal of Research in Interactive Marketing*, *12*(4), 509–523. https://doi.org/10.1108/JRIM-01-2018-0023

Homburg, C., Müller, M., & Klarmann, M. (2011). When should the customer really be king? On the optimum level of salesperson customer orientation in sales encounters. *Journal of Marketing*, *75*(2), 55–74. https://doi.org/10.1509/jm.75.2.55

Hopp, T., & Gangadharbatla, H. (2016). Novelty effects in augmented reality advertising environments: The influence of exposure time and self-efficacy. *Journal of Current Issues & Research in Advertising*, *37*(2), 113–130. https://doi.org/10.1080/10641734.2016.1171179

Huang, T.-L., & Liao, S. (2015). A model of acceptance of augmented-reality interactive technology: The moderating role of cognitive innovativeness. *Electronic Commerce Research*, *15*(2), 269–295. https://doi.org/10.1007/s10660-014-9163-2

Huré, E., Picot-Coupey, K., & Ackermann, C.-L. (2017). Understanding omni-channel shopping value: A mixed-method study. *Journal of Retailing and Consumer Services*, *39*, 314–330. https://doi.org/10.1016/j.jretconser.2017.08.011

Huynh, B., Ibrahim, A., Chang, Y. S., Hollerer, T., & O'Donovan, J. (2018). A study of situated product recommendations in augmented reality. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE. https://doi.org/10.1109/AIVR.2018.00013

Jasin, S., Sinha, A., & Uichanco, J. (2019). Omnichannel operations: Challenges, opportunities, and models. In S. Gallino & A. Moreno (Eds.), *Operations in an omnichannel world, Springer series in supply chain management* (pp. 15–34). Springer International Publishing.

Javornik, A. (2016). Augmented reality: Research agenda for studying the impact of its media characteristics on consumer behaviour. *Journal of Retailing and Consumer Services*, *30*, 252–261. https://doi.org/10.1016/j.jretconser.2016.02.004

Javornik, A., Rogers, Y., Moutinho, A. M., Freeman, R. (2016). Revealing the shopper experience of using a" magic mirror" augmented reality make-up application. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems – DIS 16*. ACM Press.

Kim, J., & Forsythe, S. (2008). Adoption of virtual try-on technology for online apparel shopping. *Journal of Interactive Marketing*, *22*(2), 45–59. https://doi.org/10.1002/dir.20113

Kim, K., Billinghurst, M., Bruder, G., Duh, H. B.-L., & Welch, G. F. (2018). Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics*, *24*(11), 2947–2962. https://doi.org/10.1109/TVCG.2018.2868591

Kocas, C. (2002). Evolution of prices in electronic markets under diffusion of price-comparison shopping. *Journal of Management Information Systems*, *19*(3), 99–119. https://doi.org/10.1080/07421222.2002.11045740

Kounavis, C. D., Kasimati, A. E., & Zamani, E. D. (2012). Enhancing the tourism experience through mobile augmented reality: Challenges and prospects. *International Journal of Engineering Business Management*, *4*, 10. https://doi.org/10.5772/51644

Kourouthanassis, P., Spinellis, D., Roussos, G., & Giaglis, G. M. (2002). Intelligent cokes and diapers: Mygrocer ubiquitous computing environment. In *First International Mobile Business Conference* (pp. 150–172).

Lackermair, G., Kailer, D., & Kanmaz, K. (2013). Importance of online product reviews from a consumer's perspective. *Advances in Economics and Business*, *1*(1), 1–5. https://doi.org/10.13189/aeb.2013.010101

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157. https://doi.org/10.1086/259131

Lee, Y., Kozar, K. A., & Larsen, K. R. (2003). The technology acceptance model: Past, present, and future. *Communications of the Association for Information Systems*, *12*(1), 50. https://doi.org/10.17705/1CAIS.01250

Ludwig, T., Hoffmann, S., Jasche, F., & Ruhrmann, M. (2020). VacuumCleanAR: Augmented reality-based self-explanatory physical artifacts. In *Proceedings of the Conference on Mensch Und Computer* (pp. 291–302).

McCluskey, M. (2022). Augmented reality is the future of online shopping.

Mora, D., Jain, S., Nalbach, O., & Werth, D. (2020). Holographic recommendations in brick-and-mortar stores.

Mystakidis, S. (2022). Metaverse. *Encyclopedia*, *2*(1), 486–497. https://doi.org/10.3390/encyclopedia2010031

Olay Skin Advisor (2020). Retrieved September 3, 2022, from https://www.olay.com/en-us/skinadvisor

Overmars, S., & Poels, K. (2015). Online product experiences: The effect of simulating stroking gestures on product understanding and the critical role of user control. *Computers in Human Behavior*, *51*, 272–284. https://doi.org/10.1016/j.chb.2015.04.033

Pantano, E. (2014). Innovation drivers in retail industry. *International Journal of Information Management*, *34*(3), 344–350. https://doi.org/10.1016/j.ijinfomgt.2014.03.002

Parekh, P., Patel, S., Patel, N., & Shah, M. (2020). Systematic review and meta-analysis of augmented reality in medicine, retail, and games. *Visual Computing for Industry, Biomedicine, and Art*, *3*(1), 1–20. https://doi.org/10.1186/s42492-020-00057-7

Park, Y., & Gretzel, U. (2010). Influence of consumers' online decision-making style on comparison shopping proneness and perceived usefulness of comparison shopping tools.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*(1), 87–131. https://doi.org/10.1146/annurev.ps.43.020192.000511

Perannagari, K. T., & Chakrabarti, S. (2019). Factors influencing acceptance of augmented reality in retail: Insights from thematic analysis. *International Journal of Retail & Distribution Management*, *48*(1), 18–34. https://doi.org/10.1108/IJRDM-02-2019-0063

Perea y Monsuwé, T., Dellaert, B. G., & de Ruyter, K. (2004). What drives consumers to shop online? a literature review. *International Journal of Service Industry Management*, *15*(1), 102–121. https://doi.org/10.1108/09564230410523358

Poushneh, A., & Vasquez-Parraga, A. Z. (2017). Discernible impact of augmented reality on retail customer's experience, satisfaction and

willingness to buy. *Journal of Retailing and Consumer Services*, *34*, 229–234. https://doi.org/10.1016/j.jretconser.2016.10.005
Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender systems – RecSys '11*. ACM Press. https://doi.org/10.1145/2043932.2043962
Ratchford, M., & Barnhart, M. (2012). Development and validation of the technology adoption propensity (tap) index. *Journal of Business Research*, *65*(8), 1209–1215. https://doi.org/10.1016/j.jbusres.2011.07.001
Rauschnabel, P. A., He, J., & Ro, Y. K. (2018). Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks. *Journal of Business Research*, *92*, 374–384. https://doi.org/10.1016/j.jbusres.2018.08.008
Rese, A., Baier, D., Geyer-Schulz, A., & Schreiber, S. (2017). How augmented reality apps are accepted by consumers: A comparative analysis using scales and opinions. *Technological Forecasting and Social Change*, *124*, 306–319. https://doi.org/10.1016/j.techfore.2016.10.010
Riar, M., Xi, N., Korbel, J. J., Zarnekow, R., & Hamari, J. (2022). Using augmented reality for shopping: A framework for AR induced consumer behavior, literature review and future agenda. *Internet Research*. Advance online publication. https://doi.org/10.1108/INTR-08-2021-0611
Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender systems handbook*. Springer.
Roitman, L., Shrager, J., & Winograd, T. (2017). A comparative analysis of augmented reality technologies and their marketability in the consumer electronics segment. *Journal of Biosensors & Bioelectronics*, *8*(1). https://doi.org/10.4172/2155-6210.1000236
Roy, S. K., Balaji, M. S., Quazi, A., & Quaddus, M. (2018). Predictors of customer acceptance of and resistance to smart technologies in the retail sector. *Journal of Retailing and Consumer Services*, *42*, 147–160. https://doi.org/10.1016/j.jretconser.2018.02.005
Schafer, J. B., Konstan, J., & Riedi, J. (1999). Recommender systems in e-commerce. In *Proceedings of the 1st ACM Conference on Electronic commerce – EC 99*. ACM Press. https://doi.org/10.1145/336992.337035
Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(6), 103–108. https://doi.org/10.9781/ijimai.2017.09.001
Smink, A. R., Frowijn, S., van Reijmersdal, E. A., van Noort, G., & Neijens, P. C. (2019). Try online before you buy: How does shopping with augmented reality affect brand responses and personal data disclosure. *Electronic Commerce Research and Applications*, *35*, 100854. https://doi.org/10.1016/j.elerap.2019.100854
Spreer, P., & Kallweit, K. (2014). Augmented reality in retail: Assessing the acceptance and potential for multimedia product presentation at the pos. *Transactions on Marketing Research*, *1*(1), 20–35. https://doi.org/10.15764/MR.2014.01002
Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the Conference on Human Factors in Computing Systems – CHI 03*. ACM Press. https://doi.org/10.1145/642611.642626
Torres-Ruiz, M., Mata, F., Zagal, R., Guzmán, G., Quintero, R., & Moreno-Ibarra, M. (2020). A recommender system to generate museum itineraries applying augmented reality and social-sensor mining techniques. *Virtual Reality*, *24*(1), 175–189. https://doi.org/10.1007/s10055-018-0366-z
Välkkynen, P., Boyer, A., Urhemaa, T., & Nieminen, R. (2011). Mobile augmented reality for retail environments. In *Proceedings of Workshop on Mobile Interaction in Retail Environments in Conjunction with MobileHCI*.
Verhagen, T., Vonkeman, C., Feldberg, F., & Verhagen, P. (2014). Present it like it is here: Creating local presence to improve online product experiences. *Computers in Human Behavior*, *39*, 270–280. https://doi.org/10.1016/j.chb.2014.07.036
Verhoef, P. C., Kannan, P., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing. *Journal of Retailing*, *91*(2), 174–181. https://doi.org/10.1016/j.jretai.2015.02.005
von Reischach, F., Guinard, D., Michahelles, F., & Fleisch, E. (2009). A mobile product recommendation system interacting with tagged products. In *2009 IEEE International Conference on Pervasive Computing and Communications*. IEEE.
Walter, F. E., Battiston, S., Yildirim, M., & Schweitzer, F. (2012). Moving recommender systems from on-line commerce to retail stores. *Information Systems and e-Business Management*, *10*(3), 367–393. https://doi.org/10.1007/s10257-011-0170-8
Wassom, B. (2014). *Augmented reality law, privacy, and ethics: Law, society, and emerging AR technologies*. Syngress.
Wierenga, B., & Van der Lans, R. (2008). *Handbook of marketing decision models* (Vol. 121). Springer.
Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41–49. https://doi.org/10.1016/j.compedu.2012.10.024
Zimmermann, R., Mora, D., Cirqueira, D., Helfert, M., Bezbradica, M., Werth, D., Weitzl, W. J., Riedl, R., & Auinger, A. (2022). Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence. *Journal of Research in Interactive Marketing*. Advance online publication. https://doi.org/10.1108/JRIM-09-2021-0237

About the authors

Jesús Omar Álvarez Márquez is a doctoral researcher at the University of Duisburg-Essen, where he is part of the Interactive Systems Group. His research focuses on group recommender systems and decision support in physical shopping scenarios.

Jürgen Ziegler is a senior professor in the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen where he directs the Interactive Systems Group. His main research interests lie in the areas of human–computer interaction, human–AI cooperation, recommender systems, information visualization, and health applications.

Appendix A. Scores obtained for functional aspects

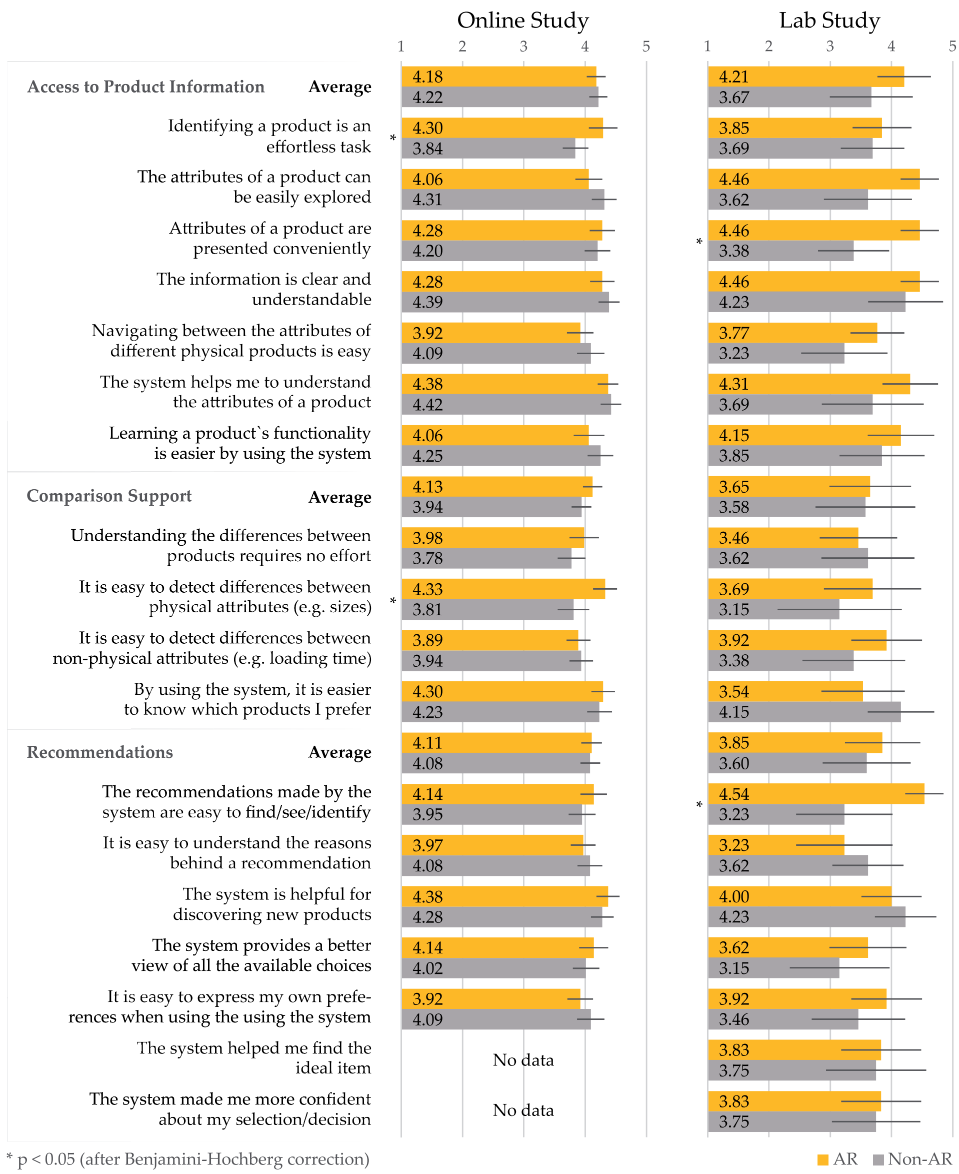


Figure A1. Means (and 95% CI) of the perceived performance of the systems, relative to the three showcased functions. Results for both online and lab studies are reported.

Appendix B. Scores obtained for functional aspects per group

	Group 1 (n = 21)				Group 2 (n = 11)				Group 3 (n = 12)				Group 4 (n = 15)				All groups (n = 59)			
	AR		Non-AR		AR		Non-AR		AR		Non-AR		AR		Non-AR		AR		Non-AR	
Access to product information	Avg	SE	Avg	SE	Avg	SE	Avg	S.E.	Avg	SE	Avg	S.E.	Avg	SE	Avg	S.E.	Avg	SE	Avg	SE
The system helps me to understand the attributes of a product.	4.43	0.148	4.52	0.148	4.36	0.203	4.64	0.152	4.25	0.250	4.25	0.179	4.27	0.15	4.33	0.187	4.34	0.089	4.44	0.085
The information is clear and understandable.	4.24	0.194	4.38	0.161	4.09	0.250	4.64	0.152	4.08	0.260	4.08	0.193	4.60	0.16	4.60	0.163	4.27	0.108	4.42	0.088
∗Attributes of a product are presented conveniently.	4.38	0.161	4.33	0.187	4.00	0.234	4.45	0.247	∗ 3.83	0.271	3.83	0.241	∗ 4.60	0.21	4.33	0.159	4.25	0.110	4.25	0.104
Identifying a product is an effortless task.	4.24	0.168	3.81	0.235	4.45	0.247	3.64	0.152	4.00	0.369	3.92	0.229	4.27	0.28	4.07	0.206	4.24	0.126	3.86	0.112
The attributes of a product can be easily explored.	4.10	0.206	4.48	0.190	4.09	0.285	4.45	0.207	3.83	0.207	4.00	0.246	4.07	0.21	4.33	0.187	4.03	0.111	4.34	0.104
∗Learning a product‘s functionality is easier by using the system.	4.29	0.171	4.43	0.177	4.18	0.296	4.18	0.226	3.42	0.379	∗ 3.67	0.256	4.07	0.27	∗ 4.53	0.192	4.03	0.135	4.25	0.110
Navigating between the attributes of different physical products is easy.	3.95	0.212	4.24	0.181	4.00	0.234	4.09	0.251	3.42	0.229	3.92	0.260	3.93	0.21	4.00	0.309	3.85	0.113	4.08	0.122
Overall	4.23	0.124	4.31	0.136	4.17	0.207	4.30	0.150	3.83	0.219	3.95	0.185	4.26	0.128	4.32	0.120	4.15	0.081	4.24	0.074
Comparison support																				
∗∗It is easy to detect differences between physical attributes.	∗∗ 4.67	0.126	4.14	0.232	4.09	0.250	3.82	0.352	∗∗/∗ 3.67	0.256	3.67	0.225	∗ 4.60	0.16	3.73	0.300	4.34	0.104	3.88	0.137
By using the system, it is easier to know which products I prefer.	4.43	0.148	4.48	0.148	4.18	0.264	4.18	0.226	3.75	0.250	4.00	0.246	4.53	0.19	4.27	0.228	4.27	0.105	4.27	0.102
Understanding the differences between products requires no effort.	4.19	0.235	4.05	0.212	3.73	0.304	3.82	0.226	3.75	0.250	3.58	0.229	3.93	0.25	3.73	0.248	3.95	0.129	3.83	0.116
∗It is easy to detect differences between non-physical attributes.	4.05	0.189	∗ 4.24	0.194	3.73	0.195	4.00	0.234	3.50	0.195	∗ 3.50	0.151	3.93	0.21	4.00	0.138	3.85	0.102	3.98	0.098
∗Overall	∗ 4.33	0.121	4.23	0.141	3.93	0.226	3.95	0.187	∗ 3.67	0.191	3.69	0.135	4.25	0.138	3.93	0.153	4.10	0.085	3.99	0.080
Recommendations																				
The system is helpful for discovering new products.	4.52	0.148	4.29	0.171	4.18	0.264	4.36	0.203	3.92	0.288	4.08	0.260	4.53	0.13	4.47	0.192	4.34	0.101	4.31	0.100
The recommendations made by the system are easy to find/see/identify.	4.29	0.171	4.14	0.221	4.00	0.357	3.73	0.273	4.00	0.213	3.83	0.241	4.13	0.22	4.07	0.206	4.14	0.112	3.98	0.117
The system provides a better view of all the available choices.	4.38	0.189	4.10	0.194	3.91	0.320	4.09	0.211	3.50	0.289	3.75	0.305	4.27	0.25	4.13	0.236	4.08	0.129	4.03	0.116
It is easy to understand the reasons behind a recommendation.	4.24	0.136	4.38	0.146	3.64	0.310	3.91	0.211	3.58	0.193	3.58	0.313	4.00	0.24	4.13	0.215	3.93	0.108	4.07	0.110
It is easy to express my own preferences when using the system.	4.00	0.169	4.29	0.197	3.82	0.296	3.82	0.182	3.50	0.261	3.83	0.271	4.20	0.20	4.20	0.296	3.92	0.112	4.08	0.122
Overall	4.29	0.112	4.24	0.134	3.91	0.283	3.98	0.169	3.70	0.204	3.82	0.242	4.23	0.15	4.20	0.180	4.08	0.090	4.09	0.089

$*p < 0.05$, $**p < 0.01$, Kruskal-Wallis test between groups, Dunn-Bonferroni pairwise comparison tests when the results were statistically significant.

Appendix C. General cluster analysis results

Table C1. Psychological traits per group.

	#	Fem.	Age	AR Kno.	Opt.	Pro.	Dep.	Vul.	TAP Score	CSO Score	RDS Score	IDS Score	RDS - IDS
					TAP Sub-scales								
Group 1	39	77%	29.08	0.08	0.05	-0.09	0.31	0.28	-0.64	1.05	0.17	0.47	0.90
Group 2	20	35%	32.40	0.37	0.23	0.55	-0.96	-0.76	2.51	-0.50	-0.26	-0.25	1.19
Group 3	24	54%	37.42	-0.44	-0.24	-0.67	0.44	0.28	-1.64	-0.73	-0.47	0.06	0.67
Group 4	33	48%	31.64	0.07	0.06	0.28	-0.25	-0.10	0.68	-0.80	0.52	-0.56	2.28
Total	116		Avg.	2.48	4.38	3.94	2.78	3.57	13.97	3.757	4.22	2.97	1.25
Outliers	11												

Psychological traits per group. From left to right: group size; percentage of females; average age; AR knowledge (on a 1–5 scale); TAP sub-scales (1–5): Optimism, Proficiency, Dependence, Vulnerability; total TAP score; Chronic Shopping Orientation (1–7); Rational and Intuitive Decision Styles (1–5); and the difference between them. Values of the psychological traits are relative to the average of the sample without outliers (last row of the table). Colored cells identify those values where a group's mean noticeably differs from the total.

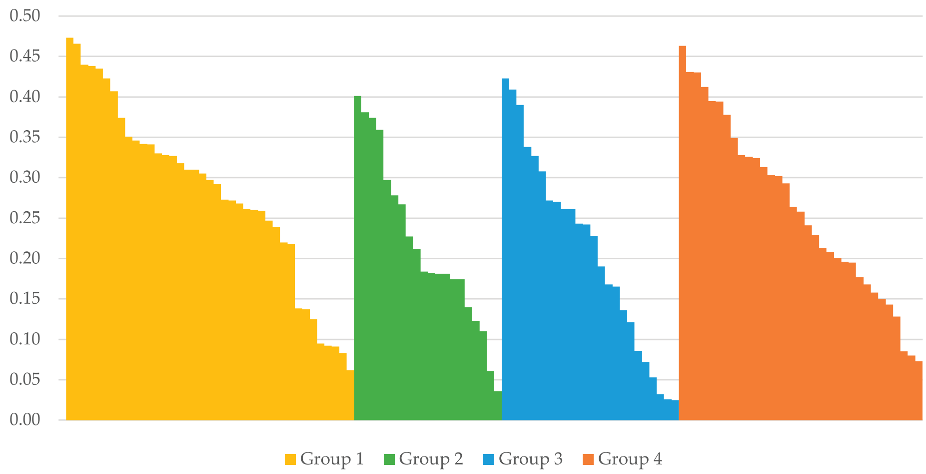


Figure C1. Silhouette scores by clusters (based on *z*-scores of psychological traits).

DOI: 10.17185/duepublico/78527
URN: urn:nbn:de:hbz:465-20230622-131227-8