JAISCR, 2023, Vol. 13, No. 2, pp. 73 – 94
10.2478/jaiscr-2023-0008

# USING COGNITIVE MODELS TO UNDERSTAND AND COUNTERACT THE EFFECT OF SELF-INDUCED BIAS ON RECOMMENDATION ALGORITHMS

Justyna Pawłowska[1,*], Klara Rydzewska[2], Adam Wierzbicki[1]

[1] *Polish-Japanese Academy of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland*

[2] *SWPS University of Social Sciences and Humanities, Chodakowska 19/31, 03-815 Warsaw, Poland*

[*] *E-mail: justyna.pawlowska@pjwstk.edu.pl*

*Submitted: 23rd August 2022; Accepted: 14th February 2023*

**Abstract**

Recommendation algorithms trained on a training set containing sub-optimal decisions may increase the likelihood of making more bad decisions in the future. We call this harmful effect self-induced bias, to emphasize that the bias is driven directly by the user's past choices. In order to better understand the nature of self-induced bias of recommendation algorithms that are used by older adults with cognitive limitations, we have used agent-based simulation. Based on state-of-the-art results in psychology of aging and cognitive science, as well as our own empirical results, we have developed a cognitive model of an e-commerce client that incorporates cognitive decision-making abilities. We have evaluated the magnitude of self-induced bias by comparing results achieved by simulated agents with and without cognitive limitations due to age. We have also proposed new recommendation algorithms designed to counteract self-induced bias. The algorithms take into account user preferences and cognitive abilities relevant to decision making. To evaluate the algorithms, we have introduced 3 benchmarks: a simple product filtering method and two types of widely used recommendation algorithms: Content-Based and Collaborative filtering. Results indicate that the new algorithms outperform benchmarks both in terms of increasing the utility of simulated agents (both old and young), and in reducing self-induced bias.

**Keywords:** recommender systems, cognitive limitations, aging, e-commerce

## 1 Introduction

The Covid-19 pandemic has transformed retail markets worldwide. One particular trend has been apparent: the acceleration of e-commerce adoption. Global e-commerce market value has risen to 26.7 trillion US dollars, while the global share of online sales in retail has risen from 14% to 19% since 2018[1]. This growth is considered fast, but it has been even faster in a particular segment of consumers: the older consumers. In some countries like the UK, online shopping has increased by threefold among older adults since the Covid-19 outbreak. A similar trend is observed in

[1] https://unctad.org/news/global-e-commerce-jumps-267-trillion-covid-19-boosts-online-sales

other countries[1]. This even faster growth of e-commerce shopping by older consumers may be the results of the superimposition of two trends: a faster adoption of e-commerce caused by the Covid-19 pandemic and the global growth of the share of people in the age group 60 plus in the population [2]. In developed countries, not only does the share of older adults in the population rise, but they are also well connected to the Internet - in the US, 73% of citizens aged over 65 had an Internet connection in 2019 [3].

The user interfaces of most contemporary e-commerce systems are usually adapted for older consumers [4], offering improved navigation and readability. Yet, this is merely scratching the surface of a complex research problem that has yet to be addressed: adapting the decision support methods used in e-commerce (and, more generally, in Web intelligence systems) for older users. The decision support methods currently in use all base on the assumption that, once a user can properly navigate the system, she or he will be able to make rational, Pareto-optimal (with respect to their individual preferences) purchasing decisions with the system's support. This assumption is negated by the well-documented effects of cognitive aging that limit the ability of older consumers to make optimal decisions [5, 6, 7].

The significance of this effect has been recognized by researchers in the HCI community, who have warned against conflating the mitigation of aging effects with improving accessibility [8]. One of the necessary adaptations concerns recommendation algorithms ubiquitously used on e-commerce platforms. In this article, we shall consider cognitive biases and limitations due to ageing [9], and the vulnerability of recommendation algorithms to a special type of bias that we call the **self-induced bias**.

**Self-induced bias of recommendation algorithms is connected to the their users' limited or deteriorating ability to make optimal decisions**. In this article, we consider internal limitations on users' decision making due to their cognitive limitations, described in detail in the literature review section. However, sub-optimal decisions could be due to external limitations, for example supply limitations (imagine a period of short supply during which users are forced to buy more expensive products, even though they prefer cheaper ones). In our case, sub-optimal decisions taken by older adults are used as a training set to create recommendations, which in turn increases the likelihood of making more bad decisions. Properly designed recommendation systems should break this vicious circle instead of replicating or magnifying it.

The study of self-induced bias due to cognitive limitations requires an interdisciplinary approach. The aim of this article is to build a bridge between computer science and psychology, and to identify important areas of joint research in the field of recommendation algorithms.

We propose a new algorithm design, aimed specifically at counteracting self-induced bias. The algorithm takes into account not only users' behaviour, but also their decision-making competency and explicitly stated preferences. To evaluate our solution, we have developed a simulation model of older consumers that is based on state-of-the-art knowledge about cognitive aging and our own empirical research results. Using the simulation model, we are able to quantify self-induced bias and to prove that the proposed algorithms have the potential to reduce it. However, the potential use of our algorithms goes beyond the support of older consumers. Our results indicate that younger consumers benefit from the proposed algorithms, as well.

The literature review on the impact of cognitive ageing on decision-making abilities and the state of art of the recommender systems addressing this issue is presented in section 2 together with a description of the authors' previous experiments used as a basis for the new model. The design and specification of the agent-based model is discussed in section 3 and is followed by the description of the simulation flow. Next, we discuss recommender systems used in the simulation and propose a new algorithm design in section 4. The results of the simulation experiments are presented in section 5. The last section concludes the article.

## 1.1 Contributions

1. Definition of new research problem concerning recommendation system trained on sub-optimal user choices: the problem of self-induced bias.

Proposal of a method to quantify self-induced bias using simulation

2. Simulator of e-commerce purchasing process supported by recommender system for clients that have cognitive limitations (code available)

3. Proposal of two recommendation algorithms that counteract self-induced bias and improve decisions of customers with cognitive limitations, outperforming optimally configured Collaborative Filtering and Content-based algorithms

# 2 Related work

## 2.1 Cognitive ageing and its impact on the decision making process

Ageing is associated with a negative changes in basic fluid cognitive abilities. Three popular ageing theories explaining the age-related changes are the Inhibition, Resources, and Speed Theories [10]. Inhibition theory explains age related deficits in cognitive performance by decreased ability to inhibit irrelevant information by older people. According to resources theory, compared to the younger adults, older adults have limited amount of cognitive resources available for allocation to a given cognitive task [11]. Finally, speed theory indicates that aging is associated with a decline in the speed with which information processing can be performed [12].

### 2.1.1 Working memory limitations

Working memory is defined as the preservation of information while simultaneously processing the same or other information[13]. Many complex everyday tasks such as decision-making, problem-solving, and the planning of goal-directed behaviors require the integration and reorganization of information from a variety of sources. There is a general consensus that working memory is impaired in older adults, although there is disagreement concerning the mechanisms involved [14]. Some results suggest that different areas are activated in young and old adults, particularly within the prefrontal cortex, indicating that younger and older adults are performing these tasks differently[15]. An understanding of age-related neurophysiological changes may help to account for these differences.

### 2.1.2 Executive control

Executive control is a multi-component construct that consists of a range of different processes that are involved in the planning, organization, coordination, implementation, and evaluation of many of our nonroutine activities [14]. It plays a significant role in virtually all aspects of cognition, allocating attentional resources among stimuli or tasks, inhibiting distracting or irrelevant information in working memory, formulating strategies for encoding and retrieval, and directing all manner of problem-solving, decision-making, and other goal-directed activities. Executive control is particularly important for novel tasks for which a set of habitual processes is not readily available.

### 2.1.3 Decision-making

While decrease in fluid cognitive abilities impact negatively performance in some decision-making task, there are also tasks where the older adults perform better than the younger ones. Bruine de Bruin et al. [16] analysing the adult age differences in decision-making competence concluded that while there is a negative relationship between age and performance on tasks which were mediated by fluid cognitive ability (Resistance to Framing, Applying Decision Rules), there is no age-related relationship or a positive age-related relationship on some tasks. The authors hypothesized that in the second type of tasks (Consistency in Risk Perception, Recognizing Age-group Social Norms, Under/Overconfidence, Resistance to Sunk Costs) older adults compensate the worsening fluid cognitive abilities with experience. They noted that although difficult to measure, the experience-related abilities like crystallized cognitive abilities, such as specialized knowledge and decision strategies, as well as emotion-related abilities, such as improved processing of affective information might improve with age.

The same research showed that although not all decision-making tasks showed age-related declines in performance, older adults perceived themselves as worse decision makers, which in turn may lead to lower motivation. This, combined with the fact that mental costs for searching increases with worsening of cognitive abilities, makes older adults put less value on informational gain and more on achieving a satisfactory decision outcome. While younger adults are often quite likely to be systematic and use maximizing decision strategies, older adults often search for information only for as long as it is sufficient. According to the selective engagement theory [6] the magnitude of this effect can be less prominent in real-life situation than in experimental setting as the older adults may be less willing to maximise effort in mentally taxing tasks. Both age-related decline in cognitive resources and motivational changes may interplay contributing to the reduction in information search among older adults.

#### 2.1.4 Strategies in multi-attribute choice

**Product comparison that is a popular feature of e-commerce platforms is a task called multi-attribute choice in psychology**. As the products are usually complex, the user needs to consider many **attributes** of the product, such as price, technical parameters, visual presentation, additional features etc. Numerous research has indicated that older adults perform worse than younger adults in such cognitively demanding experimental decision making tasks. To take into consideration all the important attributes, a person must use more complex strategies requiring thorough information processing [17, 18]. The most typical strategies identified for such tasks are: **Weighted Additive (WADD)**, **TALLY** and **Take the Best (TTB)**. Weighted Additive is an information-intensive strategy, which requires that a decision maker specifies their preferences about product attributes using weights (which can be integer values, for example, from 1 to 6). Then, during a product comparison task, each alternative product is assigned a sum of weights of the attributes for which the product has the best value among all products in the comparison (in case of a tie, all the best products add the attribute weight to their total). The product with the largest sum of weights is selected in the comparison. This strategy allows for compensation, where two attributes favoring one alternative product can counterbalance another attribute for which another alternative product has a better value. Another, yet simpler, compensatory strategy is TALLY [19], where attribute weights are ignored (or can be considered equal). For each alternative product, the total number of attributes for which the product is best in the comparison with others is counted (hence the name of the TALLY strategy). Again, the alternative with the largest total is selected.

Take The Best (TTB) is the simplest of the strategies considered here. It uses one, most important attribute. The decision is made by looking at the most important attribute and its respective values. In the case of a tie (equal values of the most important attribute among two or more products in the comparison), TTB continues the comparison using attributes of second-highest importance, and so on.

Older adults have difficulty applying decision rules that are necessary for selection of an alternative from a set of choices [20]. In multi-attribute choice tasks older adults have considerable problems with learning value of cues [21] and options [22].

### 2.2 Recommender systems for users with cognitive limitations

A comprehensive review of a psychologically informed recommender systems was conducted in [23]. Part of the recommender systems listed in the review can be classified as systems addressing the issue of cognitive limitations. The earliest adoptions of the concept include the user modelling via stereotypes [24], which although more complex is still being proposed as a solution for problems like cold-start scenarios in CF [25].

Cognitive models of memory and its limitation are being used in recommender systems in different ways. Recommender system where a cognitive model of human long-term memory is used to resemble how a human expert makes recommendations is proposed in [26]. This system can model various human memory effects such as the fan effect [2]. Bollen et al. [27] exploit positivity effects from human memory theory to investigate temporal dynamics of ratings in recommender systems. The models of losing memory over time - Ebbinghaus forgetting curve [28] is used to account for shifts in user interests in works like [29, 30, 31, 32]. Entire cognitive architectures designed to reflect how different cognitive domains work together are also being used in recommender system design and test-

[2]recognition times for a concept increases as more information is available about the concept

[3]short for adaptive control of thought-rational

ing, among which the most notable is the cognitive architecture ACT-R [3] [33]. ACT-R model was used in the context of recommender systems mainly to model the activation equation of human memory. In this model, the probability that a piece of information (i.e., a memory unit) will be activated to achieve a processing goal depends on its usefulness in the current context as well as a human's prior exposure to this information. Such a model has been used in the study of recommender systems in works: [34, 35, 36, 37, 38, 39].

However, the **differences in cognitive capacity among individuals have not been addressed even in psychologically informed recommender systems**. For example Mozer and Lindsey [40], who follow a hybrid approach that integrates collaborative filtering and computational models of forgetting, such as a variant of the above described ACT-R activation equation assume the same memory retention patterns for older and young users.

#### 2.2.1 Simulations of Recommendation Systems

When it comes to testing new solutions applicable to Recommendation Systems, experiments in a live environment with actual human users come with a high cost and risk of detrimental user experience. Moreover, few researchers have access to real-world recommender systems. Simulations have been put forward as a solution. In this approach, user feedback or decisions are usually simulated based on logged historical user data.

Most research on simulations of recommender systems is focused on predicting relevance of items to individual users and based on this information, simulating the user interaction with the recommender system (e.g., a click, a rating, dwell time, an order, etc). Initially, simulators were designed for specific datasets and specific recommendation tasks, like news article recommendation [41] or real-life retail platform and it's customers [42], which makes them difficult to use for more general research. As a solution, several simulation platforms like RecoGym [43], RecSim [44], MARS-Gym [45] and PyRecGym [46] have been developed, allowing for a varying degree of flexibility in modelling products, users, system and the interaction. Bernardi and colleagues [47] summarised the specifications of the simulators from the industry perspective.

The existence of user biases causes a challenge for off-line testing of recommender systems. This issue has been addressed in [48] with a method for debiasing the data, however there is no realistic implementation of user bias or cognitive limitations in the cited works. Our contribution to the field of recommender system simulators is incorporating cognitive limitation in the user model, based on state-of-the-art research in psychology and marketing science.

#### 2.2.2 Agent-based model of e-commerce customer using Recommendation System

Our first study of self-induced bias in recommender systems [49] was an attempt to model the behavior of e-commerce platform users (agents) with emphasis on reflecting decision-making characteristics of older adults. We have used an agent-based model to verify how cognitive deficits of e-commerce customers influence the effectiveness of collaborative filtering and content based recommender systems.

Results suggested that systems recommending items based on a larger pool of agents, like CF and the naive popularity voting model, are more effective when part of the population makes suboptimal choices due to age-related cognitive limitations. However, the simulation model used in this study had several limitations, especially because of simplified modeling of human decision making and making use of simplified product sets. In this article, we introduce a significantly improved simulation model of consumer decisions that takes into account cognitive limitations. The new simulational model is based on empirical data on user's preferences in on-line shopping, and introduces a much more comprehensive and sophisticated consumer model that incorporates crucial cognitive characteristics and decision making heuristics.

### 2.3 Bias and fairness

As noted in a comprehensive review of a literature on bias in recommended systems by Chen and colleagues [50], recent years have seen a surge of research effort on recommendation biases. However, the term is used inconsistently across papers and is used to describe several different phenomena. Researchers differentiate bias in explicit feedback data (selection and conformity biases), implicit feedback data (exposure bias and position bias), bias in model (inductive bias) and bias in results (popularity bias and unfairness).

While the self-induced bias as defined in this article matches the criteria of a general data bias [51] [4], in the context of recommender systems biases it does not fit into any of the already defined categories:

**Selection Bias** happens as users are free to choose which items to rate, so that the observed ratings are not a representative sample of all ratings. In other words, the rating data is often missing not at random. This bias in focused on the issue that certain items are not chosen to be rated, and in most cases only the particularly good or particularly bad items are reviewed. The self-induced bias is focused on the choices the user actually made.

**Conformity Bias** happens as users tend to rate similarly to the others in a group, even if doing so goes against their own judgment, making the rating values do not always signify user true preference. Self-induced bias is not related to any group or peer pressure.

**Exposure Bias** happens as users are only exposed to a part of specific items so that unobserved interactions do not always represent negative preference. Exposure bias occurs as users are only exposed to a part of items so that unobserved interactive data does not always mean negative signal. Exposure bias will mislead both the model training and evaluation. It is true that in the specific case of elderly users in e-commerce certain items might be excluded from their scope of interest for reasons unrelated to their preferences, which fits into category of exposure bias. However, the self-induced bias is a different problem that can manifest, apart from 1) lack of interaction with some items, also in 2) positive interactions with objectively non-desirable items and 3) negative interactions with objectively desirable items (eg. in our model rejecting item based on non-favourite brand). Hence although in theory the methods proposed for case 1) should help to solve part of the problems connected with self-induced bias, they definitely cannot be treated as a solution to the general issue.

**Position Bias** happens as users tend to interact with items in higher position of the recommendation list regardless of the items' actual relevance so that the interacted items might not be highly relevant. **Inductive bias** denotes the assumptions made by the model to better learn the target function and to generalize beyond training data. **Popularity Bias** Popular items are recommended even more frequently than their popularity would warrant. Position, Inductive and Popularity biases are not related to the self-induced bias.

**Unfairness**. The system systematically and unfairly discriminates against certain individuals or groups of individuals in favor others [51]. The case of older users we described in the article the self-induced bias indeed leads to recommended systems being effectively discriminatory agains users with cognitive limitations in a sense, that it consistently proposes them the less optimal items than the other groups. It is worth noting that this phenomena occurs despite the fact, that the classical recommender models we implemented do not take into account the sensitive attribute (in this case - users age). Moreover, including the sensitive attribute in a way we propose in our new systems, leads to minimising unfairness. Also the fact, that self-induced bias cannot be corrected by the methods proposed typically for reducing unfairness (like **rebalancing** [52] the data to add more observations for underrepresented groups, **regularisation** or **adversarial learning** to remove any direct or indirect information about sensitive attributes) suggest that this is a different kind of a problem than usually described in the literature. Also the methods for assuring **counterfactual fairness** [53] are not applicable, given the suboptimal recommendations are result of causally dependent characteristics of older users, such as lower working memory or different decision making strategy.

Self-induced bias occurs when users, due to their limited decision-making abilities, provide a recommendation algorithm with suboptimal decision data. For example, John, an older consumer, wants to buy a new fridge, but does not understand the technical details and does not want to spend much time considering available options. He buys an expensive model of celebrity-endorsed brand X, despite the fact that he could easily find the same quality much cheaper. The e-commerce site he

[4] systematic distortion in the sampled data that compromises its representativeness

used for the purchase remembers his choice, and when a month later John comes to replace a broken microwave, the site recommends an expensive microwave from brand X.

In summary, although self-induced bias is a type of data bias and holds some similarity to certain known subtypes of biases, the differences of it's origins and nature together with the inability of applying solutions that are effective for other bias types make self-induced bias a distinct research problem.

### 2.4 Conclusion from related work review

Work on recommender systems has attempted to take into account findings from psychology and cognitive sciences. Yet, to date there have been no attempts to address differences in cognitive capacity among individuals, or among subgroups of the user population. While researchers have developed many approaches to measure and reduce recommender systems' biases, these approaches have focused on bias due to insufficient training data for a certain user subgroup. **The problem of self-induced bias of a recommendation system has not been considered in the literature before**. It is worth noting that while we are considering self-induced bias due to cognitive aging, self-induced bias is a much more general problem. It can occur in any situation when one subgroup of users is affected by a cognitive limitation of their decision making, more strongly than the rest of the population.

## 3 Agent-based model of e-commerce customer

In this section we describe the design and parameters of a simulation model of e-commerce consumers that incorporates the effects of cognitive aging. The model is one of the main contributions of this article and can be used to study the effects of cognitive limitations on consumer purchasing decisions, satisfaction, and recommendation algorithm performance. All model assumptions and parameters are based on research in psychology of aging, consumer behavior, or our own empirical research results.

### 3.1 Model overview

The model, build using Python 3.6 language[5], consist of 3 key elements: Agents, Items and Recommender Systems. The purpose of the model is to allow an exploration of benefits and limitations of various designs of recommender systems. Agents are modelled based on older consumer's decision making characteristics described in psychological and cognitive research, as well as data collected in our experiment (see Section 2.2.2). Items are real-life products webscrapped from one of the most popular e-commerce platforms in the country. Recommender systems will be a target for detailed and realistic modeling. We also propose and test two new recommender system designs aimed specifically at dealing self-induced bias.

The model is a more advanced version of the model presented in our previous research [49]. Table 1 compares the first version of the model with the current version used for the purpose of this article. For reference, we compare our model with a well-established cognitive model - ACT-R. ACT-R is designed primarily to model cognitive processing on a lower level, hence it has a better precision in simulating memory retrieval. For the purpose of modelling customer decisions, we developed our own architecture that includes the role of emotions or brand, as well as heuristics for product comparison.

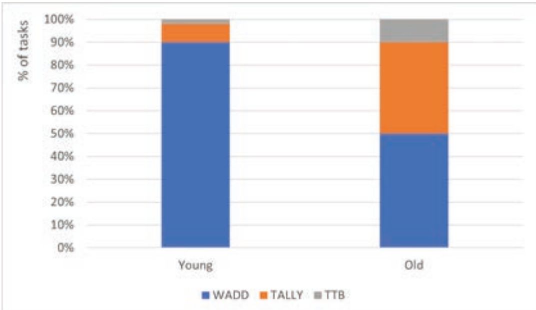


**Figure 1**. Distribution of decision strategies in multi-attribute choice problems among older and younger adults. Source: [18].

### 3.2 Empirical basis of model parameters

In order to gather empirical data for our model, we conducted an online experiment that investigated the behavior of older users in product compar-

[5]The source code is available in the repository: https://github.com/Justyna-P/Cognitive-ageing-model

**Table 1**. Comparison of simulation models of cognitive processes involved in decision making

| **Model feature** | **Basic model** [49] | **Enhanced model** | **ACT-R** [54] |
|---|---|---|---|
| Individual product preferences | YES, randomly assigned | YES, sourced from experiment | YES |
| Decision-making strategy | Not incorporated in the model | YES | YES |
| Working memory | YES | YES | YES, complex model |
| Emotion-based heuristics | YES, chosen by author | YES, calibrated based on research | Not incorporated in the model |
| Consideration sets size | Not incorporated in the model | YES | Not incorporated in the model |
| Product sets | Randomly generated | Based on real products | Not incorporated in the model |

ison tasks. The experiment involved several product comparison tasks, and obtained results confirmed the existence of age-related differences in performance among the participants. However, the details are out of scope of this article; the reader is referred to [55]. Here, let us describe relevant aspects of obtained results.

In the beginning of the study, participants filled out a personal questionnaire that asked for their age. **There were 76 users in the youngest group (aged 19-30), 80 in the middle age group (42-53) and 87 in the oldest group (65-76)**.

Participants were asked to answer questions about their personal preferences regarding product features which were later used to calibrate the simulation model used in this article. In the first questionnaire, participants rated the importance of attributes used in the product comparison tasks on a scale from 1 to 6. In another questionnaire, participants were asked how many different products they would review before deciding which one to purchase. Later, participants solved a visual pattern memory task that was successfully applied in a large internet study with participants across adult life-span [56] to measure working memory capacity.

The data collected in our empirical experiments was used to determine some of the key model parameters. **Each participant of the experiment was represented by a single Agent belonging to the same age group as the real counterpart**[6]. The user ratings of product attribute importance obtained from the first questionnaire were directly used as weights for the Agent's preferences function. The questionnaire about the number of considered products supplied our model with the size of consideration sets for each Agent - see below.

### 3.3 Model design - agents

Every agent in our model has the following properties:

1. individual product preferences - obtained from our own experiments described in section 3.2
2. a decision making strategy - randomly chosen from distribution of decision making strategies from [18]
3. working memory size - randomly chosen from distribution of age-dependent working memory size from [13]
4. susceptibility to biases (heuristics): brand or negative reviews based on research: [57, 58]

[6] The youngest and middle group both were modelled as 'young' agents

5. number of considered products (size of consideration set)- - obtained from our own experiments described in section 3.2

#### 3.3.1 Individual preferences and preference function

Every agent has a unique set of preferences, expressed as weights of each attribute of an item. The 'real' utility of an item for the agent is a weighted sum of the item's normalised attribute values weighted by the agent's preferences. For example, for the item shown in table 2, the utility for the agent would be $6*1+4*0+5*0.5+6*1+4*0+2*0.2+4*0 = 14.9$. We assume, however, that the agents do not make decisions perfectly due to cognitive limitations, hence they use simplified decision-making strategies instead of performing the full utility computation. Nevertheless, every agent has his own utility function, personalized by her preferences. This utility function allows us to measure the quality of the agent's decisions in the simulation.

#### 3.3.2 Strategies in multi-attribute choice problems

Every agent in our simulator has a decision-making strategy. Recall from Section 2.1.4 that three such strategies have been studied in research on cognitive aging: Weighted Additive (WADD), TALLY and Take The Best (TTB). These strategies have different complexities, ranging from very complex (WADD) to very simple (TTB). The WADD strategy can be considered optimal, but requires that the decision maker calculates, for each compared product, a total of the weights of attributes for which the product is best in the comparison. TALLY uses a count of these attributes for each product, while TTB only requires that the decision maker compares values of the most important attributes (unless there is a tie, in which case the next attribute in the order of importance is used).

Agents can be assigned decision strategies randomly from the distribution shown on Figure 1.The age-related proportions of decision strategies used for multi-attribute choice problems (described in 2.1.4) were taken from the experiment of Mata and colleagues [18]. When WADD was the most optimal strategy for a task, young participants were able to use it in 90% of tasks, while older participants used it in only 50% of tasks. TALLY was applied in 8% of the tasks in the younger group and in 40% of the older group, while TTB in 2% and 10%, respectively.

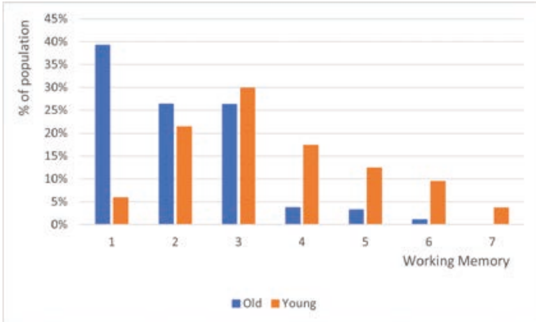


**Figure 2**. Distribution of working memory computation span (number of items) among older and younger adults. Source: [13].

#### 3.3.3 Working memory

The working memory concept (see 2.1.1) and its decrease with age play an important role in modelling fluid intelligence and its impact on the decision making process. In the model of the Agent, the working memory size parameter specifies how many attributes can be remembered and compared by the Agent at the same time. Multi-attribute choice decisions require assessing many attributes at once. A distribution of average working memory in younger (20-30) and older (60-70) adults, derived from [13], is shown on Figure 2. Agents in our simulation were assigned working memory parameter values randomly from this distribution.

#### 3.3.4 Affective Heuristics

The model so far was built under an assumption that the Agents may have a limited cognitive capacity, but their decisions are not influenced by emotions. There are multiple studies showing that emotions have a strong impact on our decision making process. Affective heuristics cause a tendency to deviate from a logical choice in a systematic way, observed repeatedly in many individuals. Older and younger adults may be impacted by affective heuristics to a varying degree. The first affective heuristic incorporated in the model is an overly strong reliance on affect-rich reviews. Research conducted by von Helversen and colleagues [57] shows that when comparing two products, one of which has a significantly better quality and av-

**Table 2**. Example of an Agent's preferences and normalized item attribute values

| attribute | A0 | A1 | A2 | A3 | A4 | A5 | A6 |
|---|---|---|---|---|---|---|---|
| importance for agent | 6 | 4 | 5 | 6 | 4 | 2 | 4 |
| item's normalised attribute value | 1 | 0 | 0.5 | 1 | 0 | 0.2 | 0 |

erage consumer rating, 18% of younger adult and 31% of older adults choose the worse product when the better one had a single highly negative, vivid, and affect-rich review. Additionally, 12% of young adults chose the worse product if it has a single enthusiastic review, which was not observed among older participants. To reflect this in the model, we have added a 'sensitive to negative review' attribute to Agents, and randomly choose 18% of younger and 31% of older Agent for which the parameter was set to 1 (for the rest of the Agents the parameter was 0). While assessing the products with negative reviews, an Agent subtracts 100 from the perceived utility meaning that the product will never be chosen over one without such a review.

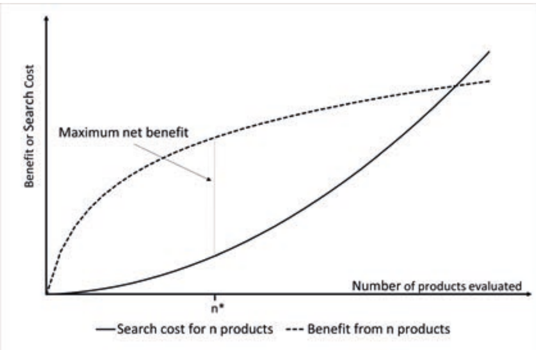


**Figure 3**. Rationality of the consideration sets. Source: [59].

A second affective heuristic present with different frequency among older and younger customers is excessive brand loyalty. Research based on car purchase data [58] has shown that even among customers not satisfied with their previous experience with a car from a given brand (satisfaction score 3/10 and below), 38% of younger and 50% of older customers still purchased the product from the same brand. We added a 'brand sensitivity' attribute to our Agent model and assigned a positive value for 38% younger and 50% older Agents. The brand to which each agent is loyal was assigned randomly among 13 different washing machine brands present in the dataset proportionally to the number of model within each brand. While assessing the product belonging to Agent's favorite brand Agent adds 100 to the perceived utility meaning that the product will always be chosen over one not belonging to the brand.

**Table 3**. Consideration set sizes in various age groups.

| Age group | average number of products considered |
|---|---|
| 1. (20-30 years) | 19.63 |
| 2. (42-52 years) | 13.45 |
| 3. (65-76 years) | 10.45 |

### 3.3.5 Consideration sets

A developed market offers a variety of products to consumers. When considering durable, complex products such as washing-machines, computers or cars, there are hundreds of available models. When consumers face a large number of alternative products, as is increasingly common in today's retail and web-based shopping environments, they typically reduce the full set of products to a smaller, more-manageable consideration set which they evaluate further[60]. This phenomenon is firmly rooted in both the experimental and prescriptive marketing literature and can be explained by a simple model assuming decreasing marginal benefit and increasing marginal search cost (see 3). Experiments[61] have shown that the average size of the consideration set are usually small – 9.3 for mobile phones, 7.8 for handheld GPSs (4.8 standard deviation), around 10 for cars. The questionnaire collected during the experiment described in Section 2.2.2 resulted in similar numbers.

### 3.4 Product sets and relative utility

There are 268 products, scrapped from one of the most popular e-commerce sites. The items are presented to the user in a random order. **Each user has one product that fits his preferences best, compared to the others**; the utility of this product for this user is the largest she can possibly achieve. **We divide the values of each agent's utility function for every product by this agent's highest utility, normalizing the utility function to** $[0,1]$.

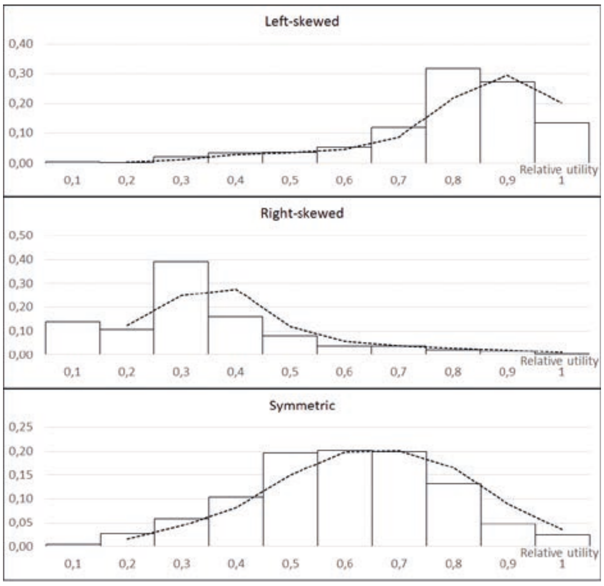


**Figure 4**. Distribution density of product's relative utility in 3 product sets

As Figure 4 shows, there are three different sets of products. First, left-skewed is based on the original products web-scrapped from an e-commerce page. The distribution of the relative utility is concentrated around the higher values - the producers compete with each other to deliver the products best suiting the client's need. To test how recommender systems would work on less developed markets, we generated product sets in which fewer products are very suited to a client's preferences: the right-skewed product set. The third product set lies between the two extremes, with a majority of products of moderate utility and less numerous items being very well or very poorly suited.

### 3.5 Simulation Flow

An agent enters the e-commerce platform with the aim to purchase a product. After entering their request into the search engine, they are presented with a set of items (P1, P2, ...Pk). The agent filters the items using minimum or maximum values of product attributes, then browses the remaining items one by one. In the beginning of the "shopping session", each agent chooses features that will be considered while making their decision. This is necessary due to limited working memory capacity; most agents are not able to simultaneously compare all the product features, so an agent with working memory capacity of 3 compares and remembers values of only 3 features most important for her. Knowing an agent's utility function and preferences, we can calculate both the "real" utility of the chosen product, that takes into account all products features, as well as "perceived" utility, that takes into account only selected features and is modified by agents biases and heuristics.

After each action, the level of the agent's motivation decreases, and when an agent's motivation level reaches 0, the agent chooses the best product seen so far (according to her "perceived" utility). Since the agent can review just a limited number of products from the entire product set, the order in which the agent browses the products is important, as only the first n product will be considered by the agent. We propose two versions of the simulation flow. In the first version, the products meeting the minimum criteria (simple product filtering) are presented to the agent in a random order. In the second version, the order in which products are presented is determined by a recommendation algorithm.

#### 3.5.1 Simple product filtering

Simple product filtering used in the first version of the simulation reflects the behaviour of a user who knows that she is unable to browse through all the products available, and does not want to see the products that do not meet chosen criteria. The assumption used in the simulation is that a user filters out all the products that for each of the features remembered by the user as important, have a value worse than the 10th percentile value among all products.

#### 3.5.2 Modeling decision support algorithms

The recommender algorithms used in the second part of the simulation determine the order in which the products are presented to the agent. **Recommendation algorithms use purchase data**

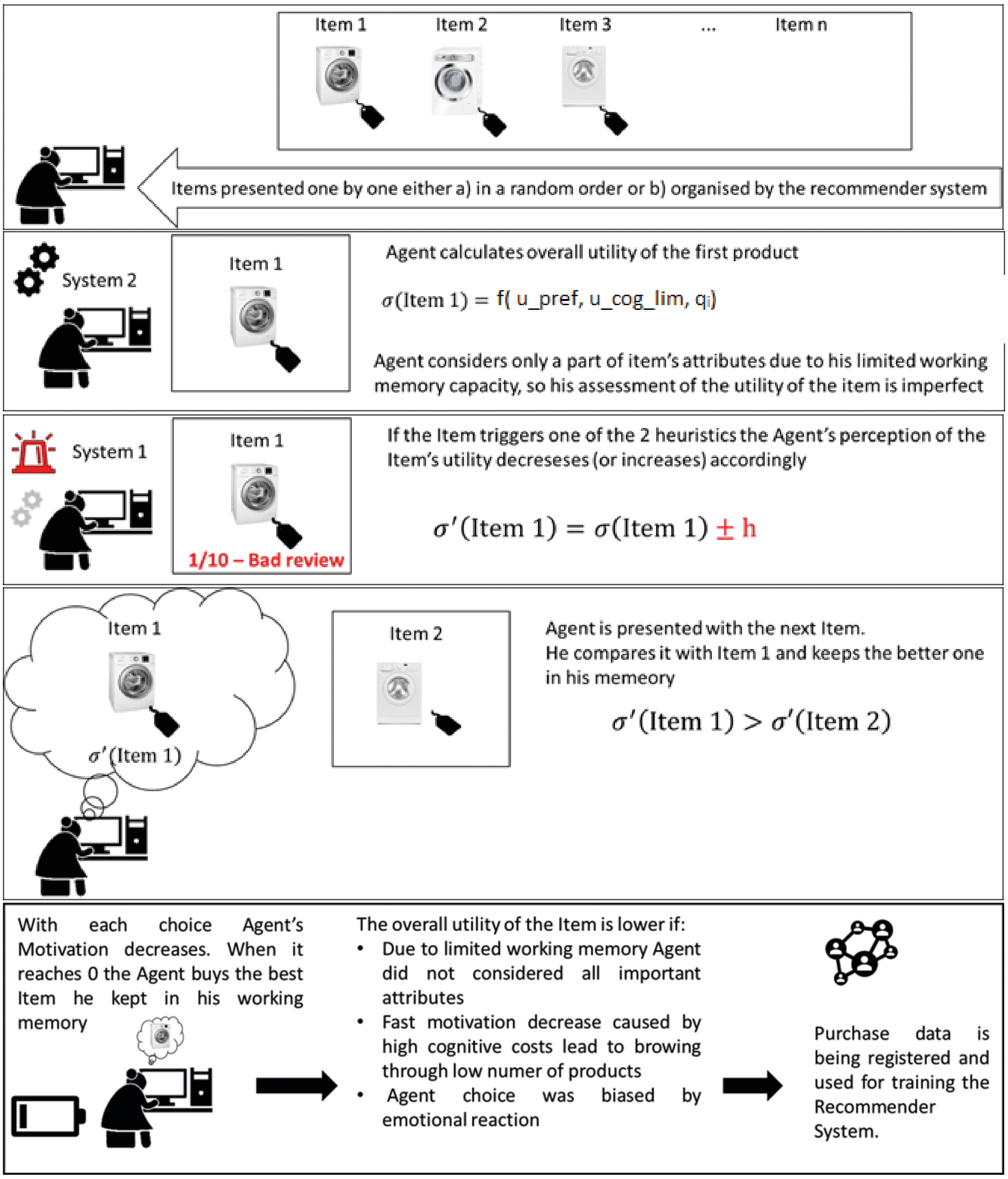


**Figure 5**. The main simulation procedure. Legend: σ (Item x) – real overall utility of the Item x for a given Agent

**from the first part of the simulation as a training set**. We simulated the experiment with the recommender algorithms on the same set of 243 agents and 268 products.

After training, an algorithm proposes the product that best matches the agent's preferences as the first displayed item, the second-best as a second item etc. In this way, the order in which an agent browses filtered products is changed by the recommendation algorithm.

# 4 Recommendation algorithms

## 4.1 Reference algorithms

### 4.1.1 Content-Based Recommendation Algorithms

Content-Based (CB) Algorithms are based on product features[7]. Such algorithms compare possible options to items a user chose in the past, and recommend most similar items [62]. Content-based algorithms are based purely on the active user's choices and do not depend on other users' data. The version of CB implemented for this research is based on matrix of item-to-item cosine similarity and the user's past purchase data. For each of the user's previously purchased items, we determined a list of 10 most similar items, then the lists were aggregated and the most frequently appearing items were recommended to the user.

### 4.1.2 Collaborative Filtering Algorithms

Collaborative Filtering (CF) techniques combine the active user's own experience and preferences with the experiences of other users. Among collaborative recommendation approaches, methods based on nearest-neighbors are the most popular [63]. This approach is based on the assumption that if user u rated several items in a similar manner to user v, then it is highly probable that user u will rate a new item the same way as user v did. For this simulation we implemented a popular latent factor model: Singular Value Decomposition (SVD). An important parameter in this approach is the number of factors $k$ to factor the user-item matrix. Higher number of factors results in higher precision, while reducing the number of factors increases the model's generalization. After running the simulation with $k$ in range (5, 10, 15, 20) we determined that k = 15 results in the highest user's utility, so this parametrisation was used for the comparison to other recommender systems.

**Algorithm 1** Proposed new algorithms

**function** RECOMMENDATION ALGORITHM (SELECTION CRITERIA)
    $u \leftarrow$ user
    $uList \leftarrow user\ list$
    $p(u) \leftarrow$ list of items purchased by user $u$
    $pref(u) \leftarrow$ list of preferences of user $u$
    $age(u) \leftarrow$ age of user $u$
    $wm(u) \leftarrow$ working memory capacity of user $u$
for each user n in uList:
    decreasingly **sort** $uList \backslash$u by the similarity of *pref(u)* to *pref(n)*
    **remove** users from *uList* based on (selection criteria)
    $topp \leftarrow$ sum of $p(u)$ for $u$ in top $k$ on *uList*
    *recList* $\leftarrow$ most frequent items in topp
    **recommend** items from *recList* to user n

**procedure** PREFERENCES TWIN
    $selectionCriteria \leftarrow age(u) =$ "old"
    **call** Recommendation algorithm (*selectionCriteria*)
**procedure** DECISION-COMPETENCY BASED
    $selectionCriteria \leftarrow wm(u) < 5$
    **call** Recommendation algorithm (*selectionCriteria*)

## 4.2 Proposed new algorithms

### 4.2.1 Preferences twin

The Preferences Twin algorithm is the first of two algorithms proposed in this article that aim to explicitly combat self-induced bias. This algorithm uses explicit user preferences and chooses a younger user with similar preferences to the older one.

The algorithm accesses the preferences database to identify the agents with the most similar preferences (using a standard k-NN algorithm [8]).

[7] The python code for the Content Based and Collaborative Filtering algorythms was based on article https://www.kaggle.com/code/gspmoreira/recommender-systems-in-python-101/notebook

[8] For the simulation k=1 was used for both proposed algorithms

Then, among these users, the algorithm selects only the "young" agents on purpose, to avoid proposing items that were chosen in a more biased and less efficient decision making process by "older" agents. The items that were purchased by the most similar, "young" agents are proposed to the agent in the simulation.

In practice, the Preferences twin algorithm would require a method of assessing user preferences. In our model, we have assumed that these preferences are expressed as weights assigned to the most important product attributes. However, there exist numerous methods of expressing user preferences, developed in decision support research. A systematic review of them [64] highlights 2 methods traditionally used for the purpose- Value Function Elicitation and Analytic Hierarchy Process and provides examples of using these methods in e-commerce space. Using a similarity in opinions between the users to improve quality of recommendation can be found in systems based on a trust-network [65].

#### 4.2.2 Decision-competency based

The Decision-competency based algorithm is the second algorithm proposed in this article. This algorithm uses explicit information about a user's decision-making abilities. As our model of a user's cognitive abilities requires a lot of input data, we have decided to focus on one of the most significant model parameters: working memory size. The Decision-competency based algorithm selects a user with similar preferences, but a working memory of at least 5.

In practice, the Decision-competency based algorithm would require not only information about a user's preferences, but also about a user's working memory size. Fortunately, there exist several well-studied and effective tests of working memory like the visual pattern span task described in Section 3.2 [56].

## 5 Simulation results

Using the simulator described in the previous section, we have conducted a number of experiments. The experiments can be classified into two categories:

1. **experiments without a recommender system**, used both for testing the sensitivity of the model to parameters' change, and for producing the training sets for the recommender algorithms
2. **experiments with a recommender system**, used for testing the effectiveness of recommender systems using different algorithm types and parameters

In this section, we first discuss the results of simulation without recommender systems.

### 5.1 Experiments without a recommender system - agent model sensitivity analysis

The first set of experiments without the recommender system was conducted to test the impact of the agent attributes on the achieved relative utility. Attributes tested were:

1. individual product preferences (based on the pilot experiment participant's answers)
2. working memory size
3. susceptibility to biases (different share of the population susceptible to biases, from 0% to 100%)
4. size of consideration set

#### 5.1.1 Experiments without a recommender system - sensitivity to product utility

The original product utility distribution was strongly *left-skewed* (see Fig 4). This is in line with expectation of the product pool on a developed market: in order to stay competitive, producers focus on delivering products fit for preferences of the majority of users. There are some outliers, products that have a low utility for an average user. This may be either comparatively less competitive products (eg. an overpriced washing machine of similar parameters as a cheaper one), or products designed to fit specific preferences (eg. very small washing machine that is not desired by an average user but is very desirable for a group of users for whom it is the only size that fits in their flat).

We have used the simulator to conduct experiments in which the utility distribution of products for different agents was *right-skewed* (a lot of products did not fit the average users' preferences, with just a few hard to find exceptions) and *symmetrical* (most of the products moderately suitable for the users).

### 5.1.2 Baseline simulation of age effects

The results of simulations without recommender systems can be used to describe the differences between "young" and "old" agents, and to introduce baseline results.

The aggregated results of simulations in which no recommender systems was used are shown on Figure 6. The x-axis shows the size of the consideration set. The y-axis shows average utility of selected products. The dotted line shows the average utility obtained by an agent who browsed respectively 1, 2 ...n items. The solid line represents older agents. The simulation is conducted three times, first on the left-skewed product set (black lines), secondly on the right-skewed product set (gray line) and lastly on the symmetrical procuct set.

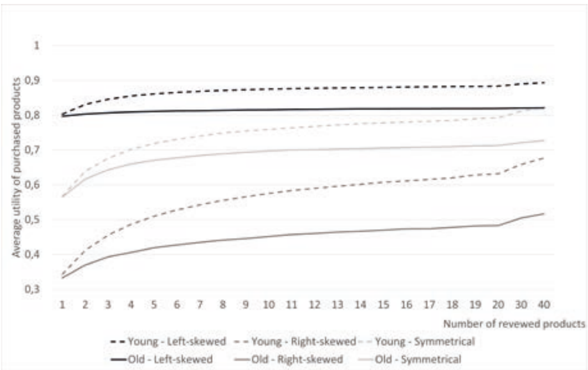


**Figure 6**. The average utility of a purchased product in a simulation without a recommender system

On Figure 6, the age difference is clear: younger agents achieve on average 5 percentage points higher utility for each number of compared products, compared to the older customers, in the left-skewed product set scenario. In the scenarios when the product set is less optimised to the customers' needs, the difference is even larger: for the symmetrical product set it is on average 6 p.p. (for every consideration set size) and if the product set is right-skewed, the difference is 11 p.p.

The age effect is even more pronounced when age-related differences in the size of the consideration sets (see Table 3) is taken into account. The achieved utility grows with the number of reviewed products for younger and older customers, and because older customers stop their search after browsing on average 10.45 items, while the younger ones consider up to 19.63 items on average, the utility of finally purchased product by the younger users is higher by 6 p.p. for the left-skewed product sets, 9 p.p. for the symmetrical product set and a staggering 18 p.p. for the right-skewed product set.

### 5.1.3 Working memory effect

The fact that the older agents have a lower working memory capacity impacts their ability to achieve a high utility of a purchase. As shown on Figure 7, both younger and older users achieve on higher average utility with an increase of their working memory capacity, although the marginal gain is diminishing. The fact that 47% of older users are not capable of simultaneously processing values of more than one attribute is consistent with findings described in [13]: the only strategy available for such users is effectively TTB.

Note that the distribution of working memory among younger users is quite broad, as well, and it is quite common for younger users to have a low working memory. This means that younger users (and younger agents in our simulation) can also make sub-optimal purchasing decisions and could also benefit from specialized support.

### 5.1.4 Impact of affective heuristics

The impact of affective heuristics on the average achieved utility is shown in Table 4. The utility achieved by older agents under influence of affective heuristics is only 1 p.p. lower than utility of the other agents. This shows that affective heuristics do not significantly harm their users, and might be a way of saving already limited cognitive resources. For younger users, the difference is more prominent, as a young agent that uses affective heuristics achieves 5 p.p. lower utility than young agents who do not use them.

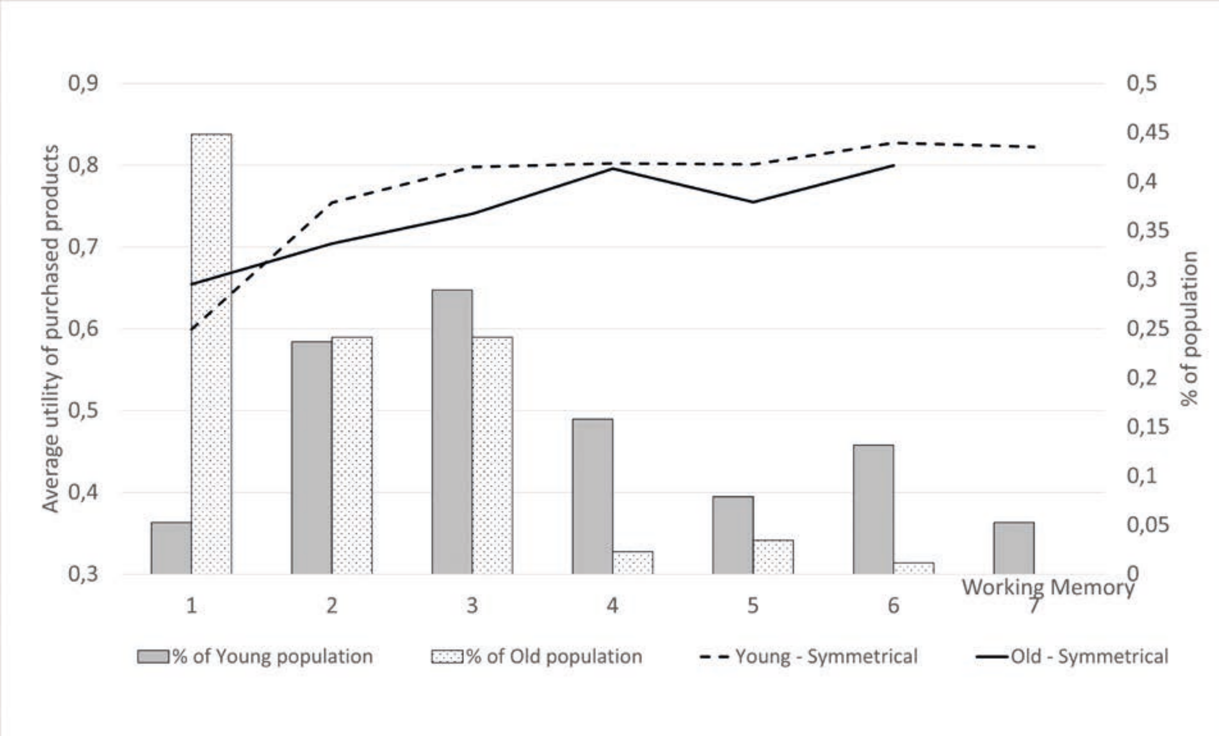


**Figure 7**. The average utility of a purchased product in a simulation without a recommender system - working memory effect

**Table 5**. The average utility of a purchased product in a simulation with a recommender system, using a left-skewed, right-skewed and symmetric distribution of product utility
*best out of all paremeters used in the simulations **difference between the utility obtained using the recommender system and the utility without the system.
All differences are statistically significant with p-values $<0.01$. The test used was Mann-Whitney U test with sample sizes $>10000$.

| Age, metric | Baseline | CB* | CF* | Preference twins | Decision competency-based |
|---|---|---|---|---|---|
| Left-skewed distribution of product utility | | | | | |
| Old, absolute | 0,82 | 0,82 | 0,84 | 0,85 | **0,86** |
| Old, relative** | NA | 0,00 | 0,02 | 0,03 | **0,04** |
| Young, absolute | 0,88 | 0,89 | 0,91 | 0,89 | **0,93** |
| Young, relative** | NA | 0,00 | **0,03** | 0,00 | **0,03** |
| Right-skewed distribution of product utility | | | | | |
| Old, absolute | 0,45 | 0,54 | 0,60 | 0,67 | **0,69** |
| Old, relative** | NA | 0,09 | 0,15 | 0,22 | **0,24** |
| Young, absolute | 0,63 | 0,68 | 0,72 | 0,77 | **0,78** |
| Young, relative** | NA | 0,05 | 0,09 | 0,14 | **0,15** |
| Symmetric distribution of product utility | | | | | |
| Old, absolute | 0,70 | 0,73 | 0,77 | 0,80 | **0,85** |
| Old, relative** | NA | 0,04 | 0,07 | 0,10 | **0,15** |
| Young, absolute | 0,79 | 0,80 | 0,86 | 0,87 | **0,89** |
| Young, relative** | NA | 0,01 | 0,07 | 0,08 | **0,09** |

**Table 4**. The average utility of a purchased product in a simulation without a recommender system - impact of affective heuristics
All differences are statistically significant with p-values <0.01. The test used was Mann-Whitney U test with sample sizes >10000.

| Age | No heuristics | Affective heuristic | Total |
|---|---|---|---|
| Old | 0,82 | 0,81 | 0,82 |
| Young | 0,91 | 0,86 | 0,88 |

## 5.2 Recommender system evaluation

We used the simulation to study the performance of classical and new recommendation algorithms designed to counteract self-induced bias. For Content-Based recommender systems, we tested the impact of the number of products used for training the algorithms. For Collaborative Filtering recommenders, the tested parameter was the number of factors used to factor the user-item matrix. The higher the number of factors, the more precise is the factorization of the original matrix. Therefore, if the model is allowed to memorize too much details of the original matrix, it may not generalize well for data it was not trained on. Reducing the number of factors increases the model generalization. Using these parameters, we have calibrated the CB and CF algorithms for optimum performance. In all subsequently discussed results achieved by CB or CS algorithms, the optimal setting of their parameters (the number of products or the number of factors, respectively) was used. The value of 10 products and 5 factors was chosen by testing values in a range from 5 to 20.

### 5.2.1 Reference recommender systems

The effectiveness of the two most popular recommender system types is presented in Table 5. The table shows only the results obtained using the best parametrisation of the CF and CB recommender algorithms.

Table 5 also shows the results of a sensitivity analysis to the distribution of product utilities. Recall that the realistic dataset of utilities of washing machines obtained from our experimental study is highly left-skewed. Since this distribution may not be typical for all applications of recommender systems (imagine a user trying to find a cheap product), we have conducted simulations with a left-skewed, right-skewed and symmetric distribution of product utility. The most prominent results have been achieved for the right-skewed and symmetric distributions.

The data shows that a well-configured CF systems is able to improve the average utility achieved by both old and young agents, respectively by 2 and 3 p.p. A similar improvement is not achieved in case of the CB system: it's usage in the best available configuration yields the same utility as achieved without the recommendation system, and some configurations of the CB algorithm (not shown in the table) are actually diminishing the utility achieved by the agents.

### 5.2.2 Proposed recommender systems

The effectiveness of the two proposed recommender algorithms is presented in Table 5. **All comparisons of average values in the table are statistically significant with p-values<0.01** (we used the Mann-Whitney U-test with sample sizes>10000).

For the left-skewed product utility distribution, the Preference twins algorithm has no effect on the utility achieved by younger agents, but improves the average utility achieved by older agents by 3 p.p. The second proposed algorithm, which takes into account not only preferences and age, but also the decision making competency, improves the utility achieved by younger agents by 3 p.p., which is comparable to the best collaborative filtering algorithm. For older users, the effectiveness of the new algorithm exceeds the effectiveness of CF algorithm and is shown to improve the utility achieved by the old users by 4 p.p.

For the right-skewed and symmetric product utility distributions, the decision-competency based algorithm clearly outperforms other algorithms. The difference between the utility of older agents when this recommendation algorithm is used and the baseline (product filtering without recommendations) is as much as 24 p.p. (for the right-skewed product utility distribution) or 15 p.p. (for the symmetric distribution). Interestingly, the decision-competency based algorithm also improves results achieved by younger adults, although by a smaller margin (15 p.p. and 9 p.p for the right-skewed and symmetric product utilities, respectively).

**Table 6**. Self-Induced Bias: Difference between absolute performance of young and old agents

| Product preference distribution | Baseline | CB* | CF* | Preference twins | Decision competency-based |
|---|---|---|---|---|---|
| Left-skewed | 0.06 | 0.07 | 0.07 | **0.04** | 0.07 |
| Right-skewed | 0.18 | 0.14 | 0.12 | 0.1 | **0.09** |
| Symmetric | 0.09 | 0.07 | 0.09 | 0.07 | **0.04** |

Note that younger users also achieve better results when the proposed algorithms are used. Both Preferences Twin and Decision-competency based algorithms outperform benchmarks for younger agents. As discussed before, this is due to the diversity of cognitive abilities (for example, working memory) among the population of younger users.

#### 5.2.3 Evaluation of self-induced bias

The self-induced bias of the studied recommender systems can be quantified by calculating the difference of absolute utilities achieved by younger and older agents, as shown on Table 6. In the first column, the difference in utilities of younger and older agents is shown when simple filtering is used, without any recommendation system. This value can be thought of as a baseline of the self-induced bias. Next columns show the self-induced bias when different recommendation systems are applied. For the case of the left-skewed distribution of utility, Content-Based and Collaborative Filtering algorithms, as well as the new proposed Decision-competency based algorithm are worse than the baseline. Only the Preference-twins algorithm manages to improve upon the baseline. For the right-skewed and symmetric datasets, CB and CF algorithms slightly improve on the baseline, but **the Decision-Competency Based algorithm manages to reduce self-induced bias by over a factor of two**.

### 5.3 Discussion and Limitations

The results described in this section are based on our simulation model. This model is an improved version of the model described in [49]. Compared to the previous version of the model, we have removed several unrealistic assumptions and limitations, such as the assumption that agents optimize an objective function - this has been replaced by the use of realistic decision strategies studied in empirical psychological research. All model parameters are derived either from our own experiments, or from most relevant psychological or marketing research. However, no matter how realistic the model, a computer simulation cannot fully replicate real human behavior.

The population of agents and items were sourced from real-life experiment participants and e-commerce platforms. However, these datasets may not reflect other possible markets (with different products or a different customer population). We have attempted to mitigate this limitation by conducting a sensitivity analysis of our results to the shape of the product utility distribution.

We have tested multiple representative versions (calibrations) of the mainstream recommender system. However, our tested versions of the Content-Based or Collaborative Filtering algorithms still may not represent the vast universe of all ways the CF and CB systems can be designed, so the results obtained from our versions may not be universal for all such recommendation systems.

The algorithms proposed in this article rely on knowledge about user preferences and user cognitive abilities. In practice, this information may be difficult to obtain. As such, the algorithms proposed in this article may be considered as concepts for future recommendation system development. Realistic algorithms could obtain user preferences using preference elicitation methods [64], and could estimate user cognitive abilities using psychological tests, such as working memory tests [56].

The results described in this section are based on our simulation model. This model is an improved version of the model described in [49]. Compared to the previous version of the model, we have removed several unrealistic assumptions and limitations, such as the assumption that agents optimize an objective function - this has been replaced by the use of realistic decision strategies studied in

empirical psychological research. All model parameters are derived either from our own experiments, or from most relevant psychological or marketing research. However, no matter how realistic the model, a computer simulation cannot fully replicate real human behavior.

The population of agents and items were sourced from real-life experiment participants and e-commerce platforms. However, these datasets may not reflect other possible markets (with different products or a different customer population). We have attempted to mitigate this limitation by conducting a sensitivity analysis of our results to the shape of the product utility distribution.

We have tested multiple representative versions (calibrations) of the mainstream recommender system. However, our tested versions of the Content-Based or Collaborative Filtering algorithms still may not represent the vast universe of all ways the CF and CB systems can be designed, so the results obtained from our versions may not be universal for all such recommendation systems.

The algorithms proposed in this article rely on knowledge about user preferences and user cognitive abilities. In practice, this information may be difficult to obtain. As such, the algorithms proposed in this article may be considered as concepts for future recommendation system development. Realistic algorithms could obtain user preferences using preference elicitation methods [64], and could estimate user cognitive abilities using psychological tests, such as working memory tests [56].

## 6 Conclusions and future research

The goal of our research has been to better understand the nature of self-induced bias of recommendation algorithms that are used by older adults with cognitive limitations. We have approached this goal by creating a simulation model of an e-commerce client that incorporates cognitive decision making abilities. This model is based on state-of-the-art results in psychology of aging and cognitive science, as well as on our own empirical results.

We have evaluated the magnitude of self-induced bias by comparing results achieved by simulated agents with and without cognitive limitations due to age. We have also proposed new recommendation algorithms designed to counteract self-induced bias. The algorithms take into account user preferences and cognitive abilities relevant to decision making.

To evaluate the algorithms, we have introduced several benchmarks: a simple product filtering method and two types of widely used recommendation algorithms: Content-Based and Collaborative filtering. Results indicate that the new algorithms outperform benchmarks both in terms of increasing the utility of simulated agents, and in reducing self-induced bias.

However, the potential of our algorithms, especially the Decision-competency based algorithm, extends beyond the combating of self-induced bias. In our simulations, younger agents have always obtained better results compared to the benchmarks, when the new algorithms were used. The reason for this lies in the differences in cognitive abilities in the population. For example, working memory distribution in the younger population has a large standard deviation, and it is quite common among younger people to have a low working memory. A practical implementation of the Decision competence based algorithm would only require a standard working memory test which could be taken online by all users of an e-commerce platform.

We believe that our research is only a first step in the study of self-induced bias in recommendation systems. More research is needed to understand this phenomenon in experimental settings. Moreover, future work can consider how the algorithms proposed in this article would work in practice, and be improved.

### Acknowledgements

This research was supported by grant 2018/29/B/HS6/02604 from the National Science Centre of Poland.

## References

[1] A. Kovalenko, Older adults shopping online: A fad or a trend?, In: The Impact of Covid-19 on E-Commerce. Proud Pen, 2020.

[2] Q. Ma, A. H. Chan, and P.-L. Teh, Bridging the digital divide for older adults via observational training: Effects of model identity from a genera-

tional perspective, Sustainability, vol. 12, 2020, p. 4555.

[3] Pew-Research, Internet/Broadband Fact Sheet, Pew Research Center, 2021.

[4] Nielsen-Norman, UX Design for Seniors (Ages 65 and older), Nielsen Norman Group, 2020.

[5] G. Sedek, P. Verhaeghen, and M. Martin, Social and motivational compensatory mechanisms for age-related cognitive decline, Psychology Press, 2012.

[6] T. M. Hess, Selective engagement of cognitive resources: Motivational influences on older adults' cognitive functioning, Persp. Psychol. Science, vol. 9, 2014, pp. 388–407.

[7] G. Sedek, T. Hess, and D. Touron, Multiple Pathways of Cognitive Aging: Motivational and Contextual Influences, Oxford University Press, 2021.

[8] B. Knowles, V. Hanson, Y. Rogers, A. M. Piper, J. Waycott, N. Davies, A. Ambe, R. N. Brewer, D. Chattopadhyay, M. Deepak-Gopinath, et al., The harm in conflating aging with accessibility, Comm. of the ACM, 2020.

[9] R. Nielek, J. Pawlowska, K. Rydzewska, and A. Wierzbicki, Adapting algorithms on the web to deal with cognitive aging, Multiple Pathways of Cognitive Aging: Motivational and Contextual Influences, 2021, p. 368.

[10] R. Cabeza, Hemispheric asymmetry reduction in older adults: the harold model., Psychology and aging, vol. 17, 2002, p. 85.

[11] D. Kahneman, Attention and effort, vol. 1063, Citeseer, 1973.

[12] J. Cerella, Age-related decline in extrafoveal letter perception, Journal of Gerontology, vol. 40, 1985, pp. 727–736.

[13] T. A. Salthouse and R. L. Babcock, Decomposing adult age differences in working memory., Developmental psychology, vol. 27, 1991, p. 763.

[14] E. L. Glisky, Changes in cognitive function in human aging, Brain aging, 2007, pp. 3–20.

[15] P. A. Reuter-Lorenz and C.-Y. C. Sylvester, The cognitive neuroscience of working memory and aging., 2005.

[16] W. Bruine de Bruin, A. M. Parker, and B. Fischhoff, Decision-making competence: More than intelligence?, Curr. Directions in Psych. Science, vol. 29, 2020, pp. 186–192.

[17] R. Mata, L. J. Schooler, and J. Rieskamp, The aging decision maker: cognitive aging and the adaptive selection of decision strategies., Psych. and aging, vol. 22, 2007, p. 796.

[18] R. Mata, B. von Helversen, and J. Rieskamp, Learning to choose: Cognitive aging and strategy selection learning in decision making., Psych. and aging, vol. 25, 2010, p. 299.

[19] G. Gigerenzer and D. G. Goldstein, Reasoning the fast and frugal way: models of bounded rationality., Psychological review, vol. 103, 1996, p. 650.

[20] T. M. Hess, T. L. Queen, and T. R. Patterson, To deliberate or not to deliberate: Interactions between age, task characteristics, and cognitive activity on decision making, Journal of Behavioral Decision Making, vol. 25, 2012, pp. 29–40.

[21] G. e. a. Chasseigne, Aging and probabilistic learning in single-and multiple-cue tasks, Experimental Aging Research, vol. 30, 2004, pp. 23–45.

[22] G. R. Samanez-Larkin, S. E. Gibbs, K. Khanna, L. Nielsen, L. L. Carstensen, and B. Knutson, Anticipation of monetary gain but not loss in healthy older adults, Nature neuroscience, vol. 10, 2007, pp. 787–791.

[23] E. Lex, D. Kowald, P. Seitlinger, T. N. T. Tran, A. Felfernig, M. Schedl, et al., Psychology-informed recommender systems, Foundations and Trends® in Information Retrieval, vol. 15, 2021, pp. 134–242.

[24] E. Rich, User modeling via stereotypes, Cognitive science, vol. 3, 1979, pp. 329–354.

[25] N. A. ALRossais and D. Kudenko, Evaluating stereotype and non-stereotype recommender systems., In: KaRS@ RecSys, 2018, pp. 23–28.

[26] M. F. Rutledge-Taylor, A. Vellino, and R. L. West, A holographic associative memory recommender system, In: 2008 Third International Conference on Digital Information Management. IEEE, 2008, pp. 87–92.

[27] D. Bollen, M. Graus, and M. C. Willemsen, Remembering the stars? effect of time on preference retrieval from memory, In: Proceedings of the sixth ACM conference on Recommender systems, 2012, pp. 217–220.

[28] H. Ebbinghaus, Memory: A contribution to experimental psychology, Annals of neurosciences, vol. 20, 2013, p. 155.

[29] H. Yu and Z. Li, A collaborative filtering method based on the forgetting curve, In: 2010 International conference on web information systems and mining, vol. 1. IEEE, 2010, pp. 183–187.

[30] L. Ren, A time-enhanced collaborative filtering approach, In: 2015 4th International Conference on Next Generation Computer and Information Technology (NGCIT). IEEE, 2015, pp. 7–10.

[31] A. Chmiel and E. Schubert, Using psychological principles of memory storage and preference to improve music recommendation systems, Leonardo Music Journal, vol. 28, 2018, pp. 77–81.

[32] Z. Yang, J. He, and S. He, A collaborative filtering method based on forgetting theory and neural item embedding, In: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE, 2019, pp. 1606–1610.

[33] J. R. Anderson, M. Matessa, and C. Lebiere, Act-r: A theory of higher level cognition and its relation to visual attention, Human–Computer Interaction, vol. 12, 1997, pp. 439–462.

[34] L. Van Maanen and J. N. Marewski, Recommender systems for literature selection: A competition between decision making and memory models, In: Proceedings of the 31st Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society, 2009, pp. 2914–2919.

[35] D. Kowald, P. Seitlinger, C. Trattner, and T. Ley, Long time no see: The probability of reusing tags as a function of frequency and recency, In: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 463–468.

[36] C. Trattner, D. Kowald, P. Seitlinger, S. Kopeinik, and T. Ley, Modeling activation processes in human memory to predict the reuse of tags, The Journal of Web Science, vol. 2, 2016.

[37] D. Kowald, P. Seitlinger, S. Kopeinik, T. Ley, and C. Trattner, Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic tag recommender, In: Mining, Modeling, and Recommending'Things' in Social Media, pp. 75–95. Springer, 2013.

[38] D. Kowald and E. Lex, The influence of frequency, recency and semantic context on the reuse of tags in social tagging systems, In: Proceedings of the 27th ACM Conference on Hypertext and Social Media, 2016, pp. 237–242.

[39] C. Stanley and M. D. Byrne, Comparing vector-based and bayesian memory models using large-scale datasets: User-generated hashtag and tag prediction on twitter and stack overflow., Psychological Methods, vol. 21, 2016, p. 542.

[40] M. C. Mozer and R. V. Lindsey, Predicting and improving memory retention: Psychological theory matters in the big data era, In: Big data in cognitive science, pp. 43–73. Psychology Press, 2016.

[41] L. Li, W. Chu, J. Langford, and R. E. Schapire, A contextual-bandit approach to personalized news article recommendation, In: Proceedings of the 19th international conference on World wide web, 2010, pp. 661–670.

[42] J.-C. Shi, Y. Yu, Q. Da, S.-Y. Chen, and A.-X. Zeng, Virtual-taobao: Virtualizing real-world online retail environment for reinforcement learning, In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 4902–4909.

[43] D. Rohde, S. Bonner, T. Dunlop, F. Vasile, and A. Karatzoglou, Recogym: A reinforcement learning environment for the problem of product recommendation in online advertising, arXiv preprint arXiv:1808.00720, 2018.

[44] E. Ie, C.-w. Hsu, M. Mladenov, V. Jain, S. Narvekar, J. Wang, R. Wu, and C. Boutilier, Recsim: A configurable simulation platform for recommender systems, arXiv preprint arXiv:1909.04847, 2019.

[45] M. R. Santana, L. C. Melo, F. H. Camargo, B. Brandão, A. Soares, R. M. Oliveira, and S. Caetano, Mars-gym: A gym framework to model, train, and evaluate recommender systems for marketplaces, In: 2020 International Conference on Data Mining Workshops (ICDMW). IEEE, 2020, pp. 189–197.

[46] B. Shi, M. G. Ozsoy, N. Hurley, B. Smyth, E. Z. Tragos, J. Geraci, and A. Lawlor, Pyrecgym: A reinforcement learning gym for recommender systems, In: Proceedings of the 13th ACM Conference on Recommender Systems, 2019, pp. 491–495.

[47] L. Bernardi, S. Batra, and C. A. Bruscantini, Simulations in recommender systems: An industry perspective, arXiv preprint arXiv:2109.06723, 2021.

[48] J. Huang, H. Oosterhuis, M. De Rijke, and H. Van Hoof, Keeping dataset biases out of the simulation: A debiased simulator for reinforcement learning based recommender systems, In: Fourteenth ACM conference on recommender systems, 2020, pp. 190–199.

[49] J. Pawlowska, R. Nielek, and A. Wierzbicki, Lost in online stores? agent-based modeling of cognitive limitations of elderly online consumers, In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Springer, 2019, pp. 204–213.

[50] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, Bias and debias in recommender system: A survey and future directions, arXiv preprint arXiv:2010.03240, 2020.

[51] A. Olteanu, C. Castillo, F. Diaz, and E. Kıcıman, Social data: Biases, methodological pitfalls, and ethical boundaries, Frontiers in Big Data, vol. 2, 2019, p. 13.

[52] M. D. Ekstrand, M. Tian, I. M. Azpiazu, J. D. Ekstrand, O. Anuyah, D. McNeill, and M. S. Pera, All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness, In: Conference on fairness, accountability and transparency. PMLR, 2018, pp. 172–186.

[53] M. J. Kusner, J. Loftus, C. Russell, and R. Silva, Counterfactual fairness, Advances in neural information processing systems, vol. 30, 2017.

[54] C. Dimov, P. H. Khader, J. N. Marewski, and T. Pachur, How to model the neurocognitive dynamics of decision making: A methodological primer with act-r, Behavior research methods, vol. 52, 2020, pp. 857–880.

[55] K. Rydzewska, J. Pawłowska, R. Nielek, A. Wierzbicki, and G. Sedek, Cognitive limitations of older e-commerce customers in product comparison tasks, In: IFIP Conference on Human-Computer Interaction. Springer, 2021, pp. 646–656.

[56] R. H. Logie and E. A. Maylor, An internet study of prospective memory across adulthood., Psychology and aging, vol. 24, 2009, p. 767.

[57] B. Von Helversen, K. Abramczuk, W. Kopeć, and R. Nielek, Influence of consumer reviews on online purchasing decisions in older and younger adults, Decision Support Systems, vol. 113, 2018, pp. 1–10.

[58] R. Lambert-Pandraud, G. Laurent, and E. Lapersonne, Repeat purchasing of new automobiles by older consumers: empirical evidence and interpretations, Journal of Marketing, vol. 69, 2005, pp. 97–113.

[59] J. R. Hauser, Consideration-set heuristics, Journal of Business Research, vol. 67, 2014, pp. 1688–1699.

[60] J. R. Hauser, O. Toubia, T. Evgeniou, R. Befurt, and D. Dzyabura, Disjunctions of conjunctions, cognitive simplicity, and consideration sets, Journal of Marketing Research, vol. 47, 2010, pp. 485–496.

[61] M. Ding, J. R. Hauser, S. Dong, D. Dzyabura, Z. Yang, S. Chenting, and S. P. Gaskin, Unstructured direct elicitation of decision rules, Journal of Marketing Research, vol. 48, 2011, pp. 116–127.

[62] P. Lops, M. De Gemmis, and G. Semeraro, Content-based recommender systems: State of the art and trends, Recommender systems handbook, 2011, pp. 73–105.

[63] C. Desrosiers and G. Karypis, A comprehensive survey of neighborhood-based recommendation methods, Recommender systems handbook, 2011, pp. 107–144.

[64] L. Chen and P. Pu, Survey of preference elicitation methods, Technical report, 2004.

[65] R. B. Nozari and H. Koohi, Novel implicit-trust-network-based recommendation methodology, Expert Systems with Applications, vol. 186, 2021, p. 115709.

**Justyna Pawłowska-Bebel** is a Ph.D candidate at Polish-Japanese Academy of Information Technology in Warsaw, Poland, currently completing Interdisciplinary program in Computer Science & Psychology in cooperation with SWPS University of Social Sciences and Humanities. Her research combines technological aspects of recommender systems in e-commerce space with psychological perspective of cognitive aging and decision-making processes.
https://orcid.org/0000-0002-2070-6268

**Klara Rydzewska** received her PhD in Psychology from the SWPS University of Social Sciences and Humanities in Warsaw, Poland, in 2017. Currently, she works at the Department of New Technologies in Social Applications and the Interdisciplinary Center for Applied Cognitive Studies (ICACS) at the SWPS University in Warsaw. Her main research interests include cognitive aging, decision making, and applications of new technologies. She published articles in Psychology and Aging, Behavioral and Brain Sciences, and chapters in an edited book of Oxford University Press. She presented invited papers in Switzerland, Germany, United States, and Japan.
https://orcid.org/0000-0002-5683-6320

**Adam Wierzbicki** is a scientist working in the area of Social Informatics, an area of informatics that aims to design information systems and algorithms while taking into account their social and psychological impact, as well the reciprocal impact of human behavior on information systems. Prof. Wierzbicki has been a pioneer of research on Web content credibility. In 2013-2016, he has led the Reconcile project that researched methods of Web content credibility evaluation. He is an author of the book “Web content credibility”. Currently, he is involved in research on the credibility of medical Web content. He also co-leads a research project that designs e-commerce platforms and recommendation algorithms that deal with cognitive aging. Dr Wierzbicki is the Steering Committee Chair of the International Conference on Social Informatics (SocInfo), and a Senior Member of ACM.
https://orcid.org/0000-0003-0075-7030