Repeating my Workouts or Exploring new Activities? A Longitudinal Micro-Randomized User Study for Physical Activity Recommender Systems

Ine Coppens
imec - WAVES - Ghent University
Ghent, Belgium
Ine.Coppens@UGent.be

Toon De Pessemier
imec - WAVES - Ghent University
Ghent, Belgium
Toon.DePessemier@UGent.be

Luc Martens
imec - WAVES - Ghent University
Ghent, Belgium
Luc1.Martens@UGent.be

ABSTRACT

While repeating activities can create healthy habits, exploring new physical activities is also important to increase health benefits and prevent boredom. Following habit formation and variety-seeking behavior theories, this study investigates the difference between repetition and exploration of physical activities in health recommender systems. An eight-week Micro-Randomized Trial is conducted in which 11 physically inactive adults receive personalized activity recommendations that are either (a) a repetition or exploration, (b) connected to a specific location, duration, or neither, and (c) accompanied by videos or Points-of-Interest, or not. Analyses of the 187 submitted activity recommendations suggest that the inactive participants prefer exploration, as exploration recommendations were submitted the most, had significantly larger star ratings and durations, and are dependent on moderators (b) and (c). More specifically, exploration for workouts received the highest star ratings and motivation. To our knowledge, this study is the first to investigate repetition and exploration for physical activities, contributing to effective recommender algorithms for healthy behavior change.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; *Empirical studies in ubiquitous and mobile computing*; • **Information systems** → **Recommender systems**; **Personalization**; • **Applied computing** → *Psychology*; *Consumer health.*

KEYWORDS

behavior change, physical activity, habit formation, variety-seeking, decision-making, motivation

1 INTRODUCTION

We all repeat our preferred behaviors: relistening to our most-liked songs, revisiting our favorite restaurants, and rebuying our go-to foods at grocery stores. The reconsumption of these items can be explained by the status quo bias, in which people lean towards maintaining their current or previous decision, such as buying the same brands or staying in the same job [31, 32]. However, reconsumption of beloved songs [10], or other hedonic items, such as books, movies, or places [31], is different than re-engaging in healthy behaviors for healthy eating or physical activity (PA), because of the additional effort and motivation they require. Depending on the complexity of the behavior, repetition of behaviors make them more efficient and automatic, until they develop into a habit [23].

Nonetheless, repetition can lead to satiation and boredom [2]. Instead of repeating what they already know, people also experience variety-seeking behavior, in which they prefer exploring new items [42]. People's decision-making to maximize enjoyment is based on finding the balance between the satiation from repetition, and stimulation from variety [34]. To support people in their decision-making processes, Recommender Systems (RSs) can model user preferences and generate suggestions to explore new items that fit with their preferences [29, p vii], and thereby help them make decisions for healthy behavior change [18]. Although RSs and human decision-making are closely related, most RS research mainly focuses on the techniques and algorithms [7]. There should be more focus on the decision-making processes of the users, and whether this is supported by the RS [5, 7].

Incorporating variety and exploration in the PA recommendations is also important for people's health. The World Health Organization (WHO) recommends engaging in 150-300 minutes of moderate-intensity aerobic PA per week, and incorporating two days of muscle-strengthening activities per week for additional health benefits, because the different types of PA and the combination of aerobic and strength-promoting exercise provide favorable health outcomes [35, 39]. For example, if a person normally only engages in walking activities, even if these activities are highly preferred by the user, higher intensity or strength activities should be proposed as well.

As such, both repetition and exploration of PA behavior seem important for PA recommendations. There are different RS algorithms, and each algorithm produces a different list of recommended items [9]. For this reason, the RS developer should thoroughly decide on which RS approach and properties to implement, as their effect on user experience depends on the domain [29, p 570]. We argue that the RS algorithm should be adapted to people's decision-making processes for repeating a habit PA or exploring a new activity [7]. However, this decision-making between repetition and exploration has not been investigated in the domain of health RSs for PA, to the best of our knowledge.

In this preliminary study, we investigate whether repeating or exploring personalized PA recommendations in a mobile RS results in higher star ratings, motivation, and PA duration. The user study follows the experimental design of a Micro-Randomized Trial (MRT) that, similar to within-subject experiments, provides a more informative experiment because there cannot be a biased split of subjects into multiple groups, such as in between-subject studies [29, p 555]. Conducting our MRT study over eight weeks, this research contributes to the need of more longitudinal research to study the balance between repetition and variety [34], and to RS research in the PA domain, which requires more effort and healthy behavior change than the typical RS domains, such as video, music, and e-commerce.

1.1 Research questions

This paper is focused on the following main research question: *RQ1: Do repetition or exploration items for PA recommendations have the best effects on star rating feedback, momentary motivation to execute the PA, and PA duration?*, as illustrated in Figure 1. Based on the habit formation process of Lally et al. [23], we hypothesize that repeating PAs will increase automaticity and cause users to choose these PA items more often because they are becoming a habit over time. We expect this to result in higher star rating feedback, motivation, and PA duration, because PA habits are associated with autonomous motivation [17].

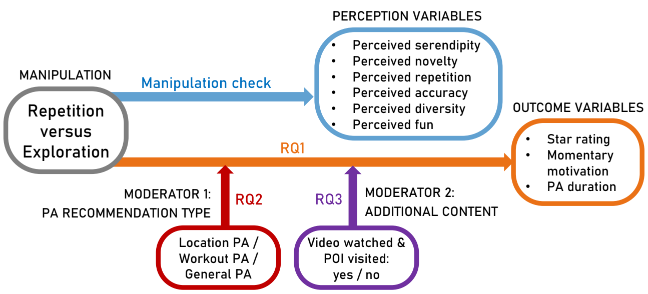


Figure 1: This conceptual framework of the study shows that a manipulation check firstly verifies whether the manipulation succeeded, followed by investigating the main effect of RQ1 and the interaction effects of RQ2 and RQ3.

When creating the PA dataset, we noticed that PAs are more heterogeneous in content than other RSs' items, such as movies or music. Some PAs can only be performed at a specific location, but are not necessarily connected to a specific duration (e.g., bowling), while others can be recommended with a specific duration at various locations (e.g., a 30-minute yoga session), and others are not attached to a location, nor duration (e.g., taking the stairs instead of the elevator). In this way, we distinguish three main categories of PA recommendations, as illustrated in Table 1. Following this idea, our second research question is: *RQ2: Does the effect of repetition versus exploration depend on the PA recommendation type of location, workout, or general?* We hypothesize that the type of PA recommendation will be a moderator, which is a third variable that alters the effect of the independent variable on the outcome variable [26], and can be observed as an *interaction effect* [15]. More specifically, we expect a preference for repetition in general and workout PAs because people can do these at any location in their daily life, making them more likely to be adopted as a habit [23].

As depicted in Table 1, location PAs can be extended with Point-of-Interest (POI) suggestions, and workout PAs with YouTube videos. As we investigate whether adding these two types of content will moderate the main effect, we formulate our third research question: *RQ3: Does the effect of repetition versus exploration depend on whether the PA recommendation also contains a POI location or YouTube video link?* We hypothesize that people are more motivated to explore a new PA when a POI or video suggestion is provided because of the additional guidance and inspiration.

To verify whether our manipulation of a repetition versus an exploration recommendation succeeded, we implemented several subjective manipulation checks using six additional questions as feedback [16]: the *perceived serendipity* and *novelty* based on [29, p 587-589], the *perceived accuracy*, *diversity*, and *fun* based on the subjective metrics of [22], and our own variable *perceived repetition* to check whether they already did the PA before.

2 RELATED WORK

Previous work investigated the repetition of consumption items in various domains, such as music relistening [10, 28, 36], POIs revisiting [8], and repeat consumptions in general [2]. On the contrary, exploration of new music taste in an RS was investigated in [25]. Research in the field of RSs also investigated Next Basket Recommendations (NBRs) that balance between repeat items and explore items, such as NBRs for groceries [3, 24]. Compared to NBRs, our user study presents both repetition and exploration items at the same time in randomized positions in an MRT study design.

Our research is applied in the health domain for PAs, which require more effort than listening to music or buying groceries. Several preceding studies have confirmed the association of PA habits and PA behavior repetition [11, 17, 38]. To generate new PA habits, Dogangün et al. [12] implemented if-then plans to provide individual recommendations. Arguing that most PA interventions still adopt a "one-size-fits-all" approach, Yfantidou et al. [41] identified strategies across population segments to facilitate personalization. Other studies have integrated personalization in PA interventions with contextualization [21], using personalized gamified feedback [33], by matching with personality traits for improved well-being [19], or by finding similar profiles among hypertensive patients using collaborative filtering to find suitable PAs that helped control blood pressure for those profiles [13]. In our study, PA recommendations are personalized with a content-based RS algorithm and an adaptation algorithm to users' context and current PA level.

Other MRT studies in the health domain found that a gamified team competition intervention (vs. no competition) significantly increases daily PA [37], that automatic tailoring of conversation topics (vs. users choosing their own topic) in a virtual health coach leads to equal user engagement [4], and that delivering an activity suggestion (vs. no suggestion) increases step count [21]. In our MRT, three levels of randomization are applied: (a) repetition vs. exploration, (b) location PA vs. general PA vs. workout PA, and (c) showing a POI/YouTube link vs. not showing it.

3 METHODS

To present the PA recommendations to the user, an Android smartphone app was developed, as illustrated in Figure 2(a). The RS algorithm uses a PA dataset, based on the Compendium of Physical Activities [1], which was extended with the PA recommendation type, as each item in the dataset belongs to exactly one type, following our idea in Table 1. Additionally, the location PA items were manually assigned with a Google Maps POI search query, and the workout PA items with a YouTube video query. Lastly, 24 binary

Table 1: Three types of PA recommendations can be distinguished, depending on whether they require a specific location to which a Point-of-Interest (POI) can be connected, a specific duration to which a YouTube video can be linked, or neither.

PA recommendation type	fixed duration	fixed location	additional content	example
location PA		X	POI	a swimming pool
workout PA	X		YouTube video	a 10-minute ab workout
general PA			nothing	stand up during phone calls

attributes were manually assigned to each PA item to describe their content for the content-based RS, such as *aerobic, flexibility, balance, indoors/outdoors*, and *alone/with buddy*.

This content-based RS algorithm is applied to generate personalized PA suggestions by matching the attributes of the user's previous PA consumption (i.e., the user profile) with the attributes of the items from the PA dataset [29, p 251]. By requesting the user's star rating feedback and whether the PA was executed indoors/outdoors and alone/with a buddy (Figure 2(b)), the RS can model the user's preferences and link this to similar items using the cosine similarity in subsequent PA recommendations [29, p 256]. The recommendations are also adjusted to the user's context and profile (e.g., the current weather, mood, remaining daylight, and available material), adapted the user's current PA level, and gradually increasing in duration following the WHO recommendations for increased health benefits [39].

At every refresh time, the RS generates six PA recommendations, one for each of the six combinations of the 2 (RQ1: repetition/exploration) x 3 (RQ2: location/workout/general) randomizations. Repetition items are created by calculating content-based recommendations only on the items that the user already submitted in the eight-week study, while exploration items are calculated only on PAs that were never submitted before by the user. As shown in an example randomization in Figure 2(a), this random assignment of the recommendations' position is our implementation of the MRT study design. It is micro-randomized for every participant for every delivery time [40], while providing all six possibilities at any time, and eliminating the effect of position bias in the list. For RQ3, the MRT is implemented by randomly toggling the video button (for workout PAs) or the POI button (for location PAs) on or off, as shown at the top of the screen in Figure 2(a). The video link is created by appending the suggested personalized duration to its query from our PA dataset (e.g., "home exercises 5 minutes"), while the POI link is only its query (e.g., "library"). When a user clicks on the video or POI button, YouTube or Google Maps opens respectively with the corresponding search query, providing multiple workout videos and POIs nearby. Out of these, the user can select their own choice.

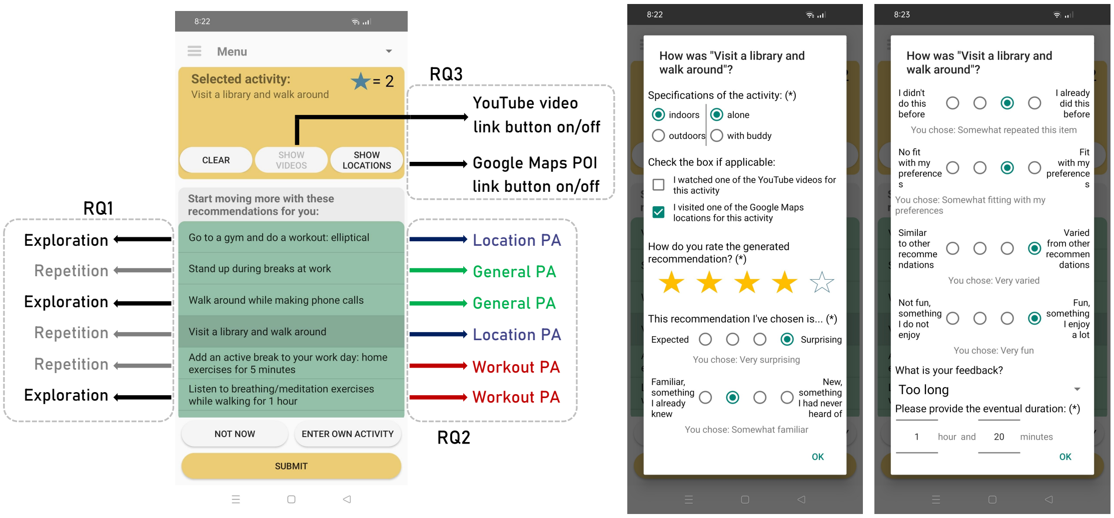
The target group of our real-life user study are healthy adults who have less than 150 minutes of moderate-to-vigorous-intensity PA per week, and were recruited via the Sona platform of Ghent University and Facebook groups for paid studies in Ghent. They install the Android app on their own smartphone, after which their eligibility in the study is checked by providing their age, their initial weekly amount of moderate-to-vigorous PA with the European Health Interview Survey - Physical Activity Questionnaire (EHIS-PAQ) [14], and their health and fitness assessment with the Revised Physical Activity Readiness Questionnaire [6]. If eligible, they use the app for eight weeks in their daily life for a maximum incentive of 30 EUR. They are informed that their incentive amount is not dependent on how many PA recommendations they engage in, but on all interactions with the app, such as submitting their own PAs or reasons why now is not a good time for PA (the buttons at the bottom of Figure 2(a)). Instead, they are rewarded with stars connected to every PA, depending on its effort and duration (at the top of Figure 2(a)). The study received ethical approval from the Ethical Committee of the Faculty of Psychology and Educational Sciences of Ghent University (https://www.ugent.be/pp/en/research/ec) on August 22, 2023 (reference number: 2023-061A).

Statistical analyses are conducted using Generalized Estimating Equations (GEEs) in SPSS Statistics v. 29 with the user ID as subject variable, an unstructured working correlation matrix [27], and effect sizes using Cramer's V [20]. The GEEs' results are reported with their Estimated Marginal Means (EMMs), and the statistically significant interaction effects with corresponding interaction plots [15]. The outcome metrics are requested in a feedback screen when submitting a recommended item: star rating feedback (measured with five stars, as shown in Figure 2(b)), the eventual duration of the PA (shown at the bottom of Figure 2(c)), the momentary motivation for the PA (measured on a 4-point Likert scale in a separate screen shown after the feedback screens), and the six perception questions of the manipulation check (also with 4-point Likert scales, as depicted in Figures 2(b) and 2(c)).

4 RESULTS AND DISCUSSION

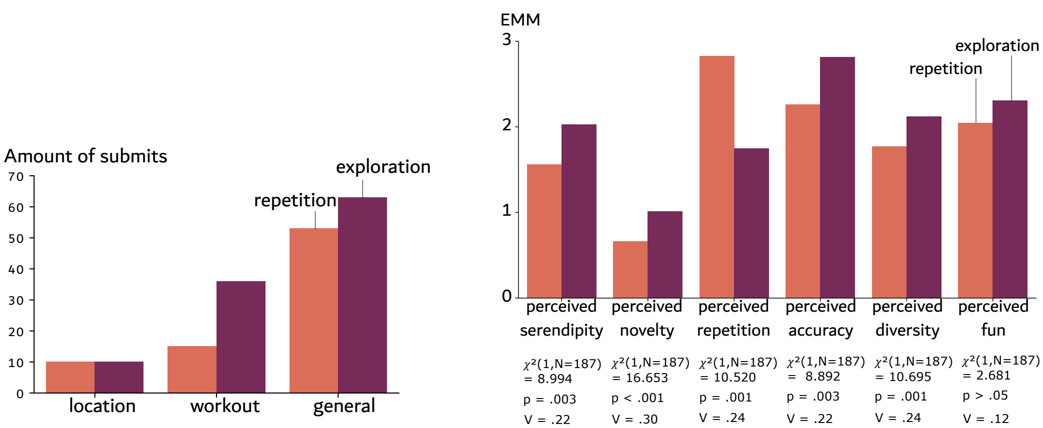
22 participants installed the app (100% in the age group of 18 - 44 years), of which 11 finished the whole eight-week study (initial weekly moderate-to-vigorous PA minutes measured with EHIS-PAQ [14]: M = 54.0, SD = 55.9, min = 0, max = 150). In total, those 11 participants submitted 187 PAs that were recommended to them (amount of submitted recommendations: M = 17.0, SD = 13.8, min = 1, max = 52). As the amount of submits greatly vary between participants, there is a possibility that few participants highly influence the results, but the GEE model takes into account these independent clusters (i.e., the participants) of dependent observations (i.e., their submits) [27]. An overview of the amount of submits, distinguished between repetition versus exploration, and PA recommendation type, is presented in Figure 3(a). This bar chart shows that PAs at a location were submitted the least, which could be explained by the higher effort to go to the location (e.g., a park). Because general PAs (e.g., a stand-up meeting) are generally lower in intensity than workouts (e.g., a dance workout), their overall lower effort could explain why general PAs were chosen the most.

Manipulation checks: Six separate GEE analyses were conducted on the manipulation check variables, of which the test results and EMMs are provided in Figure 3(b). Perceived serendipity,



(a) (b) (c)

Figure 2: In the main screen of the app, three types of randomizations can be distinguished, corresponding to each of the research questions (a). When a user wants to submit a PA recommendation item, feedback on the location, company, POI visit/video watching, star rating, manipulation check variables, and eventual duration are requested in a feedback screen (b and c).



(a) (b)

Figure 3: Of the 187 submitted PA recommendations, general PAs were submitted the most, followed by workout PAs, for which the the exploration variant was chosen more than twice as much (a). The GEEs' EMMs, χ^2 results, *p*-values, and Cramer's *V* of the manipulation check variables show that the manipulation succeeded as the participants perceived the repetition recommendations as significantly more repeating, and the exploration items as significantly more surprising, new, accurate, and diverse (b).

novelty, and diversity are significantly higher for exploration items, suggesting that users perceived the items as more surprising, new, and diverse, respectively, which was expected because these had not been recommended before. The participants also perceived the repetition items as significantly more repeating. This confirms that our manipulation succeeded and people actually repeated those PAs. Perceived accuracy was significantly higher for the exploration items, which suggests that participants found that these items fit better with their preferences. We did not find a statistically significant difference for perceived fun, which could be explained because people are more likely to choose fun items anyway, regardless of their type.

RQ1: The GEE resulted in a significant main effect on star rating ($\chi^2(1, N{=}187) = 8.746$, $p = .003$, Cramer's $V = .22$), indicating that participants rated the exploration items (EMM = 4.49 on five stars) higher than the repetition items (EMM = 4.19 on five stars). Additionally, PA duration was significantly higher for exploration items (EMM = 19.87 minutes) compared to repetition items (EMM = 11.90 minutes) (χ^2(1, N=187) = 4.042, p = .044, Cramer's V = .15). As our target group are people who initially did not attain the 150-minute weekly minimum, and because they submitted more exploration items and rated them higher on star rating and perceived accuracy, our results suggest that exploring new PAs might be preferred by physically inactive people. However, we did not find a significant main effect on motivation (χ^2(1, N=187) = 2.257, p > .05, Cramer's V = .11), which is not consistent with Hawlader et al. [17] who found that autonomous motivation is associated with PA habits.

RQ2: Conducting the GEE with PA recommendation type as moderator, results in a significant interaction effect on star rating feedback (χ^2(2, N=187) = 10.417, p = .005, Cramer's V = .17), which is illustrated in Figure 4(a). Together with the higher submit amount, these results suggest that inactive people prefer exploration for workouts. This does not support our hypothesis and could be explained by people's variety-seeking behavior: workouts, which are at higher intensities, require more user effort, and repeating them can lead to boredom [2, 42]. However, we do not know for certain that an exploration item is fully new to the user, as we only know that this item was new during the study. There is also a significant interaction effect on momentary motivation (χ^2(2, N=187) = 31.075, p < .001, Cramer's V = .29). On average, people were more motivated for repetition of location PAs and exploration of workouts, as depicted in Figure 4(b). The lower motivation for repetition of workouts can be explained because health interventions that encourage repetition can activate the healthy choice without requiring willpower [30], suggesting that motivation for the workout can be low to be chosen and executed, once it became an automated habit. Another significant interaction effect was found on PA duration (χ^2(2, N=187) = 6.011, p = .05, Cramer's V = .13) and is depicted in Figure 4(c). This line graph shows that repetition of location PAs resulted in longer durations. However, the amount of location PA submits is limited, thwarting reliable analyses. Lastly, there are limited differences for the general PAs across all three outcome variables. This could be explained because our general PAs require the least effort (no fixed duration or location, and lower intensities), resulting in equal preferences to explore or repeat them.

RQ3: For the submitted workout PAs (N=51), there is a significant interaction effect of the added YouTube video on star rating (χ^2(1, N=51) = 4.139, p = .042, Cramer's V = .28) illustrated in Figure 4(d), and on PA duration (χ^2(1, N=51) = 8.574, p = .003, Cramer's V = .41) in Figure 4(e). These results suggest that adding a video with instructions can increase the star rating for exploring new workouts, but increases repetition workouts' duration more. This might indicate that people could prefer exploration of new workouts at first with a short duration, but are motivated to repeat the same one with longer durations once they found a suitable one. We did not find a significant interaction effect on motivation (χ^2(1, N=51) = 3.471, p > .05, Cramer's V = .26). As no participants indicated to have visited a POI for the location PAs, which could be explained because only few location PAs were submitted (N=20) and because the POI button was only toggled on ±50% of the time in the MRT design, no analyses can be conducted for this moderator.

5 CONCLUSION AND FUTURE WORK

Situated in habit formation and variety-seeking behavior theories, this study investigates whether repeating or exploring PAs is preferred in an RS that personalizes PAs. We defined a distinction between three types of PA recommendations because they differ in whether their duration or location is fixed, resulting in the possible addition of a video or POI. Following the WHO guidelines [39] on PA duration for health benefits, an eight-week user study with 11 physically inactive participants (<150 minutes PA/week) investigates their star rating feedback, motivation, and PA duration.

The manipulation check confirmed that repetition recommendations were perceived as significantly more repeating, and exploration items as significantly more surprising, new, diverse, and accurate. Our results suggest that physically inactive adults prefer exploration of PAs, as these were submitted the most and resulted in significantly higher star ratings and durations. The results also show that the three outcome variables depend on the moderators. More specifically, exploring workouts had higher motivation and star ratings, which might be boosted by adding a YouTube video.

Due to small sample size, there is limited data about location PAs and the addition of videos and POIs. A larger participant pool will be recruited in a subsequent study to further investigate these additions. As continuously recommending new PA items is not possible once all items have been suggested, future work should also investigate the long-term feasibility of PA exploration with longer user studies. We also suggest to integrate the level of habit strength and behavior automaticity. In this way, an RS can define the moment at which a newly explored activity turns into a habit, and suggest increasingly longer PA durations for repeating workouts. Future researchers could also integrate a metric that defines a personalized balance between repetition and exploration, which can adapt the algorithm's settings and user model to generate the most optimal health behavior for that user's current automaticity and preferences.

ACKNOWLEDGMENTS

This research received funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme.

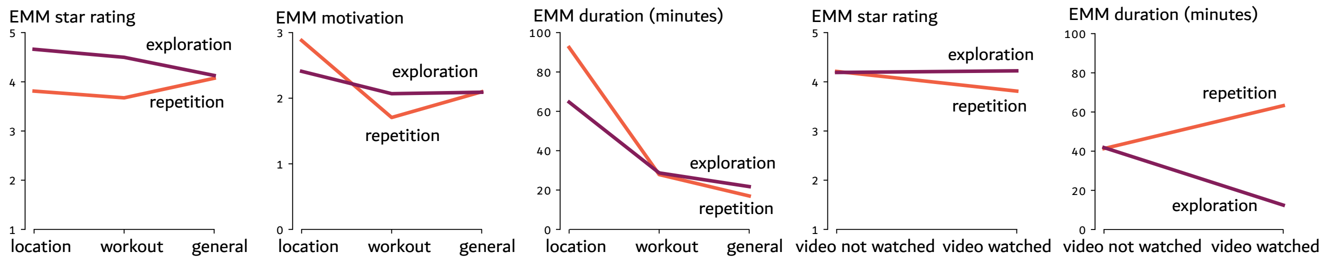
REFERENCES

[1] Barbara E. Ainsworth, William L. Haskell, Stephen D. Herrmann, Nathanael Meckes, David R. Bassett, Catrine Tudor-Locke, Jennifer L. Greer, Jesse Vezina, Melicia C. Whitt-Glover, and Arthur S. Leon. 2011. 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values. *Medicine & Science in Sports & Exercise* 43, 8 (Aug. 2011), 1575–1581. https://doi.org/10.1249/MSS.0b013e31821ece12

[2] Ashton Anderson, Ravi Kumar, Andrew Tomkins, and Sergei Vassilvitskii. 2014. The dynamics of repeat consumption. In *Proceedings of the 23rd international conference on World wide web*. ACM, Seoul Korea, 419–430. https://doi.org/10.1145/2566486.2568018

[3] Mozhdeh Ariannezhad, Sami Jullien, Ming Li, Min Fang, Sebastian Schelter, and Maarten De Rijke. 2022. ReCANet: A Repeat Consumption-Aware Neural Network for Next Basket Recommendation in Grocery Shopping. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Madrid Spain, 1240–1250. https://doi.org/10.1145/3477495.3531708

[4] Tessa Beinema, Harm Op Den Akker, Marian Hurmuz, Stephanie Jansen-Kosterink, and Hermie Hermens. 2022. Automatic topic selection for long-term interaction with embodied conversational agents in health coaching: A micro-randomized trial. *Internet Interventions* 27 (March 2022), 100502. https://doi.org/10.1016/j.invent.2022.100502



(a) (b) (c) (d) (e)

Figure 4: The interaction plots illustrate the significant interaction effects of the PA recommendation type moderator on the star rating (a), momentary motivation (b), and PA duration (c), and of the video watched moderator on star rating (d) and PA duration (e).

[5] André Calero Valdez, Martina Ziefle, and Katrien Verbert. 2016. HCI for Recommender Systems: the Past, the Present and the Future. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, Boston Massachusetts USA, 123–126. https://doi.org/10.1145/2959100.2959158

[6] Bradley J. Cardinal, Jacqueline Esters, and Marita K. Cardinal. 1996. Evaluation of the Revised Physical Activity Readiness Questionnaire in older adults:. *Medicine & Science in Sports & Exercise* 28, 4 (April 1996), 468–472. https://doi.org/10.1097/00005768-199604000-00011

[7] Li Chen, Marco De Gemmis, Alexander Felfernig, Pasquale Lops, Francesco Ricci, and Giovanni Semeraro. 2013. Human Decision Making and Recommender Systems. *ACM Transactions on Interactive Intelligent Systems* 3, 3 (Oct. 2013), 1–7. https://doi.org/10.1145/2533670.2533675

[8] Zhilong Chen, Hancheng Cao, Huangdong Wang, Fengli Xu, Vassilis Kostakos, and Yong Li. 2020. Will You Come Back / Check-in Again?: Understanding Characteristics Leading to Urban Revisitation and Re-check-in. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (Sept. 2020), 1–27. https://doi.org/10.1145/3411812

[9] Kei Long Cheung, Dilara Durusu, Xincheng Sui, and Hein De Vries. 2019. How recommender systems could support and enhance computer-tailored digital health programs: A scoping review. *DIGITAL HEALTH* 5 (Jan. 2019), 205520761882472. https://doi.org/10.1177/2055207618824727

[10] Frederick Conrad, Jason Corey, Samantha Goldstein, Joseph Ostrow, and Michael Sadowsky. 2019. Extreme re-listening: Songs people love . . . and continue to love. *Psychology of Music* 47, 2 (March 2019), 158–172. https://doi.org/10.1177/0305735617751050

[11] Gert-Jan De Bruijn, Stef P.J. Kremers, Amika Singh, Bas Van Den Putte, and Willem Van Mechelen. 2009. Adult Active Transportation. *American Journal of Preventive Medicine* 36, 3 (March 2009), 189–194. https://doi.org/10.1016/j.amepre.2008.10.019

[12] Aysegül Dogangün, Michael Schwarz, Katharina Kloppenborg, and Robert Le. 2017. An Approach to Improve Physical Activity by Generating Individual Implementation Intentions. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, Bratislava Slovakia, 370–375. https://doi.org/10.1145/3099023.3099101

[13] Luciano Rodrigo Ferretto, Ericles Andrei Bellei, Daiana Biduski, Luiz Carlos Pereira Bin, Mirella Moura Moro, Cristiano Roberto Cervi, and Ana Carolina Bertoletti De Marchi. 2020. A Physical Activity Recommender System for Patients With Arterial Hypertension. *IEEE Access* 8 (2020), 61656–61664. https://doi.org/10.1109/ACCESS.2020.2983564

[14] Jonas D. Finger, Jean Tafforeau, Lydia Gisle, Leila Oja, Thomas Ziese, Juergen Thelen, Gert B. M. Mensink, and Cornelia Lange. 2015. Development of the European Health Interview Survey - Physical Activity Questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. *Archives of Public Health* 73, 1 (Dec. 2015), 59. https://doi.org/10.1186/s13690-015-0110-z

[15] Sara Garofalo, Sara Giovagnoli, Matteo Orsoni, Francesca Starita, and Mariagrazia Benassi. 2022. Interaction effect: Are you doing the right thing? *PLOS ONE* 17, 7 (July 2022), e0271668. https://doi.org/10.1371/journal.pone.0271668

[16] David J. Hauser, Phoebe C. Ellsworth, and Richard Gonzalez. 2018. Are Manipulation Checks Necessary? *Frontiers in Psychology* 9 (June 2018), 998. https://doi.org/10.3389/fpsyg.2018.00998

[17] Mohammad Delwer Hossain Hawlader, Nusrat-E Mozid, Shakila Sharmin, Imran Hossain Monju, Sanjana Binte Ahmed, Wharesha Sarker, Mohammad Ashraful Amin, Shirin Shahadat Jhumur, and Koustuv Dalal. 2023. The art of forming habits: applying habit theory in changing physical activity behaviour. *Journal of Public Health* 31, 12 (Dec. 2023), 2045–2057. https://doi.org/10.1007/s10389-022-01766-4

[18] Katja Herrmanny and Helma Torkamaan. 2021. Towards a User Integration Framework for Personal Health Decision Support and Recommender Systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, Utrecht Netherlands, 65–76. https://doi.org/10.1145/3450613.3456816

[19] Mohammed Khwaja, Miquel Ferrer, Jesus Omana Iglesias, A. Aldo Faisal, and Aleksandar Matic. 2019. Aligning daily activities with personality: towards a recommender system for improving wellbeing. In *Proceedings of the 13th ACM Conference on Recommender Systems*. ACM, Copenhagen Denmark, 368–372. https://doi.org/10.1145/3298689.3347020

[20] Hae-Young Kim. 2017. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics* 42, 2 (2017), 152. https://doi.org/10.5395/rde.2017.42.2.152

[21] Predrag Klasnja, Shawna Smith, Nicholas J Seewald, Andy Lee, Kelly Hall, Brook Luers, Eric B Hekler, and Susan A Murphy. 2019. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. *Annals of Behavioral Medicine* 53, 6 (May 2019), 573–582. https://doi.org/10.1093/abm/kay067

[22] Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (Oct. 2012), 441–504. https://doi.org/10.1007/s11257-011-9118-4

[23] Phillippa Lally, Cornelia H. M. Van Jaarsveld, Henry W. W. Potts, and Jane Wardle. 2010. How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology* 40, 6 (Oct. 2010), 998–1009. https://doi.org/10.1002/ejsp.674

[24] Ming Li, Sami Jullien, Mozhdeh Ariannezhad, and Maarten De Rijke. 2023. A Next Basket Recommendation Reality Check. *ACM Transactions on Information Systems* 41, 4 (Oct. 2023), 1–29. https://doi.org/10.1145/3587153

[25] Yu Liang and Martijn C. Willemsen. 2019. Personalized Recommendations for Music Genre Exploration. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, Larnaca Cyprus, 276–284. https://doi.org/10.1145/3320435.3320455

[26] Sang-June Park and Youjae Yi. 2022. Assessing moderator effects, main effects, and simple effects without collinearity problems in moderated regression models. *Journal of Business Research* 145 (June 2022), 905–919. https://doi.org/10.1016/j.jbusres.2022.03.018

[27] Stano Pekár and Marek Brabec. 2018. Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology* 124, 2 (Feb. 2018), 86–93. https://doi.org/10.1111/eth.12713

[28] Markus Reiter-Haas, Emilia Parada-Cabaleiro, Markus Schedl, Elham Motamedi, Marko Tkalcic, and Elisabeth Lex. 2021. Predicting Music Relistening Behavior Using the ACT-R Framework. In *Fifteenth ACM Conference on Recommender Systems*. ACM, Amsterdam Netherlands, 702–707. https://doi.org/10.1145/3460231.3478846

[29] Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). 2022. *Recommender Systems Handbook*. Springer US, New York, NY. https://doi.org/10.1007/978-1-0716-2197-4

[30] Alexander J. Rothman, Paschal Sheeran, and Wendy Wood. 2009. Reflective and Automatic Processes in the Initiation and Maintenance of Dietary Change. *Annals of Behavioral Medicine* 38, S1 (Dec. 2009), 4–17. https://doi.org/10.1007/s12160-009-9118-3

[31] Cristel Antonia Russell and Sidney J. Levy. 2012. The Temporal and Focal Dynamics of Volitional Reconsumption: A Phenomenological Investigation of Repeated Hedonic Experiences. *Journal of Consumer Research* 39, 2 (Aug. 2012), 341–359. https://doi.org/10.1086/662996

[32] William Samuelson and Richard Zeckhauser. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* 1, 1 (March 1988), 7–59. https://doi.org/10.1007/BF00055564

[33] Hanna Schäfer, Joachim Bachner, Sebastian Pretscher, Georg Groh, and Yolanda Demetriou. 2018. Study on Motivating Physical Activity in Children with Personalized Gamified Feedback. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, Singapore Singapore, 221–226. https://doi.org/10.1145/3213586.3225227

[34] Julio Sevilla, Joy Lu, and Barbara E. Kahn. 2019. Variety Seeking, Satiation, and Maximizing Enjoyment Over Time. *Journal of Consumer Psychology* 29, 1 (Jan. 2019), 89–103. https://doi.org/10.1002/jcpy.1068

[35] Emmanuel Stamatakis, I Min Lee, Jason Bennie, Jonathan Freeston, Mark Hamer, Gary O'Donovan, Ding Ding, Adrian Bauman, and Yorgi Mavros. 2018. Does Strength-Promoting Exercise Confer Unique Health Benefits? A Pooled Analysis of Data on 11 Population Cohorts With All-Cause, Cancer, and Cardiovascular Mortality Endpoints. *American Journal of Epidemiology* 187, 5 (May 2018), 1102–1112. https://doi.org/10.1093/aje/kwx345

[36] Kosetsu Tsukuda and Masataka Goto. 2020. Explainable Recommendation for Repeat Consumption. In *Fourteenth ACM Conference on Recommender Systems*. ACM, Virtual Event Brazil, 462–467. https://doi.org/10.1145/3383313.3412230

[37] Jitao Wang, Yu Fang, Elena Frank, Maureen A. Walton, Margit Burmeister, Ambuj Tewari, Walter Dempsey, Timothy NeCamp, Srijan Sen, and Zhenke Wu. 2023. Effectiveness of gamified team competition as mHealth intervention for medical interns: a cluster micro-randomized trial. *npj Digital Medicine* 6, 1 (Jan. 2023), 4. https://doi.org/10.1038/s41746-022-00746-y

[38] Susanne Weyland, Emily Finne, Janina Krell-Roesch, and Darko Jekauc. 2020. (How) Does Affect Influence the Formation of Habits in Exercise? *Frontiers in Psychology* 11 (Oct. 2020), 578108. https://doi.org/10.3389/fpsyg.2020.578108

[39] World Health Organization. 2020. *WHO guidelines on physical activity and sedentary behaviour.* http://www.ncbi.nlm.nih.gov/books/NBK566045/ OCLC: 1237095892.

[40] Jing Xu, Xiaoxi Yan, Caroline Figueroa, Joseph Jay Williams, and Bibhas Chakraborty. 2023. A flexible micro-randomized trial design and sample size considerations. *Statistical Methods in Medical Research* 32, 9 (Sept. 2023), 1766–1783. https://doi.org/10.1177/09622802231188513

[41] Sofia Yfantidou, Pavlos Sermpezis, and Athena Vakali. 2022. 12 Years of Self-tracking for Promoting Physical Activity from a User Diversity Perspective: Taking Stock & Thinking Ahead. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, Barcelona Spain, 211–221. https://doi.org/10.1145/3511047.3538029

[42] Yuan Zhang. 2022. Variety-Seeking Behavior in Consumption: A Literature Review and Future Research Directions. *Frontiers in Psychology* 13 (June 2022), 874444. https://doi.org/10.3389/fpsyg.2022.874444