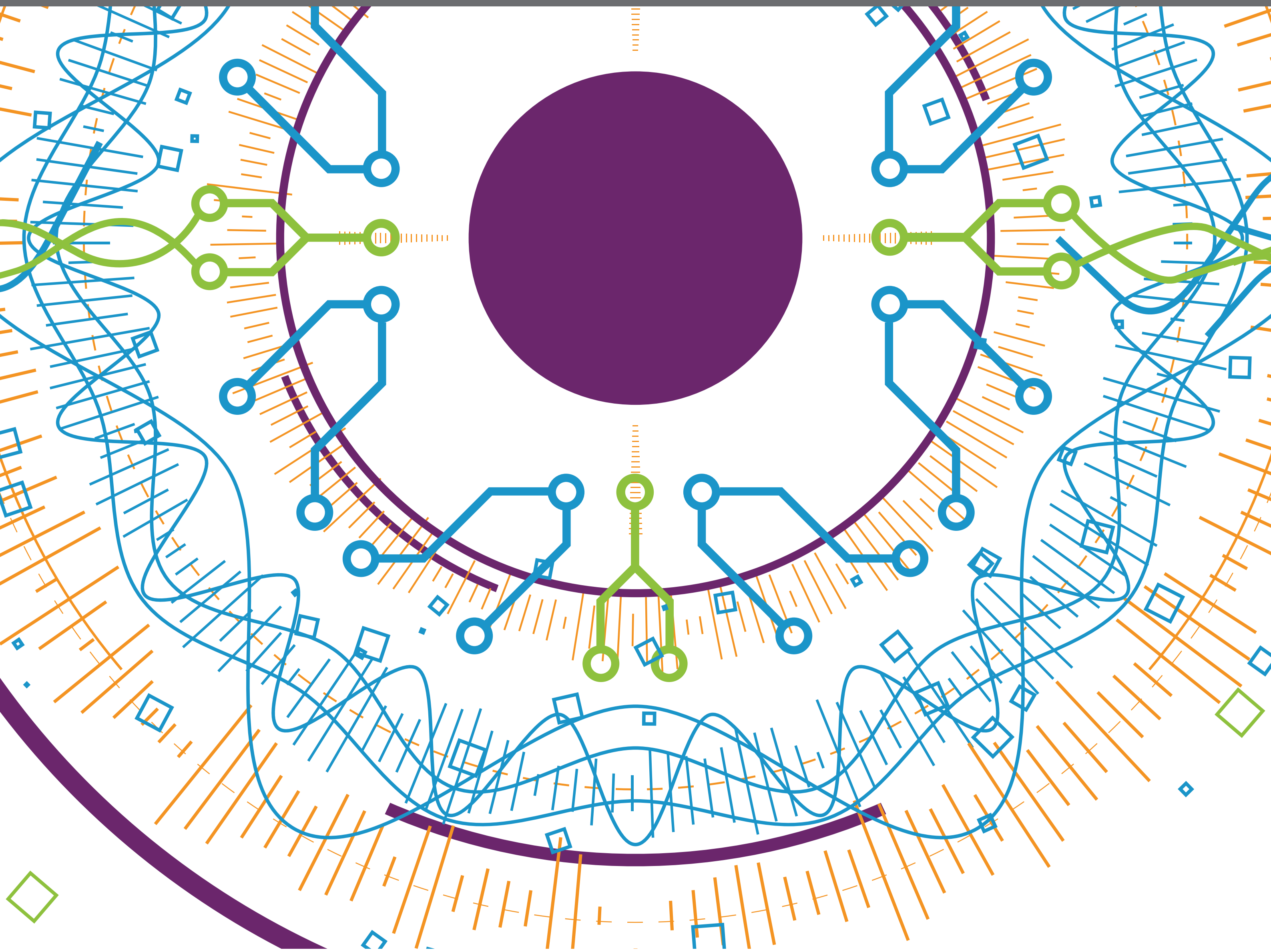
ARTICLE COLLECTION ON HUMAN ASPECTS IN ADAPTIVE AND PERSONALIZED INTERACTIVE ENVIRONMENTS (HAAPIE)

EDITED BY: Vania Gatseva Dimitrova, Panagiotis Germanakos and Styliani Kleanthous

PUBLISHED IN: Frontiers in Artificial Intelligence



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-375-0
DOI 10.3389/978-2-88966-375-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ARTICLE COLLECTION ON HUMAN ASPECTS IN ADAPTIVE AND PERSONALIZED INTERACTIVE ENVIRONMENTS (HAAPIE)

Topic Editors:
Vania Gatseva Dimitrova, University of Leeds, United Kingdom
Panagiotis Germanakos, SAP SE, Germany
Styliani Kleanthous, Open University of Cyprus, Cyprus

Citation: Dimitrova, V. G., Germanakos, P., Kleanthous, S., eds. (2021). Article Collection on Human Aspects in Adaptive and Personalized Interactive Environments (HAAPIE). Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-375-0

Table of Contents

frontiers in Artificial Intelligence

EDITORIAL
published: 27 November 2020
doi: 10.3389/frai.2020.606068

Editorial: Article Collection on the Human Aspects in Adaptive and Personalized Interactive Environments

Panagiotis Germanakos[1,2]*, *Vania Gatseva Dimitrova*[3]* and *Styliani Kleanthous*[4]*

[1]SAP SE, Walldorf, Germany, [2]Department of Computer Science, University of Cyprus, Nicosia, Cyprus, [3]School of Computing, University of Leeds, Leeds, United Kingdom, [4]Cyprus Center for Algorithmic Transparency, Open University of Cyprus, Latsia, Cyprus

Keywords: human-computer interaction, computational intelligence, human factors, personalization, adaptation, user model, algorithm, usability & user experience

Editorial on the Research Topic

Article Collection on the Human Aspects in Adaptive and Personalized Interactive Environments

OPEN ACCESS

Edited and reviewed by:
Julita Vassileva,
University of Saskatchewan, Canada

Correspondence:
Panagiotis Germanakos
panagiotis.germanakos@sap.com
Vania Gatseva Dimitrova
V.G.Dimitrova@leeds.ac.uk
Styliani Kleanthous
styliani.kleanthous@gmail.com

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *14 September 2020*
Accepted: *14 October 2020*
Published: *27 November 2020*

Citation:
Germanakos P, Dimitrova VG and Kleanthous S (2020) Editorial: Article Collection on the Human Aspects in Adaptive and Personalized Interactive Environments. Front. Artif. Intell. 3:606068. doi: 10.3389/frai.2020.606068

The rapid technological advancement and increasing availability of big data in recent years, has transformed computational systems to more dynamic, multidimensional digital communication environments that present highly complex and uncertain information flows, unfamiliar scenarios, situation-specific use cases, and multi-purpose interactions. Such a reality brings more prominently intelligent technologies to the center of attention for providing alternative insights, adaptive interventions, personalized conditions, and smart solutions to the benefit of the unique user. In principle, a vast body of research addresses adaptation and personalization based on ordinary user characteristics (e.g., role, experience, knowledge, interests) or related contextual aspects (e.g., displays, connectivity, processing power). Researchers, professionals, and practitioners in the broader scientific areas of Adaptive Hypermedia, Web Personalization, and User Modeling, have determined numerous user aspects that demonstrate an unquestionable positive influence to the content, functionality, and interactions offered by adaptive and personalized systems in various application fields (e.g., modeling the user preferences and interests for increasing the accuracy of recommender systems, or modeling knowledge and skills in educational hypermedia systems for an enhanced learning experience). Building on this premise, the extraction and use of deeper psychological constructs, values, and abilities (attributes that define individuals, e.g., cognition, intelligence, emotions, personality, expertise), may also fundamentally advance the role of computer-mediated environments that encompass human-computer interactions.

Acknowledging that human-computer interactions are essentially executed on a cognitive level (i.e., users may engage into actions that involve learning, problem solving and decision making), it is of paramount importance to scrutinize and coordinate individual traits and differences throughout the whole design and development process of current practices. Human factors may be exploited during the definition and implementation of the user models or to be regarded as an integrated intelligent component of a system producing smart user interfaces and interaction paradigms. The expected outcomes may offer to users a rich user experience and enhanced usability during the execution of their activities advancing the overall quality of computational systems, services and applications. Nevertheless, considering the multi-dimensional nature of human factors as well as the complexity of the data structures and content meta-characteristics, we could recognize that this is not an easy task. The modeling of individual characteristics and the formulation of respective rules that would guide optimal personalization environments, conditions, and functionality in various contexts and application domains is a long and cumbersome iterative process. In recent years, such individual differences have been extensively explored and utilized in adaptation and personalization systems, yielding in some cases mixed outcomes whereby in others show a significant improvement on the personalization of the user experiences. Thus, there is an indisputable need for a fundamental shift in our understanding of individual differences which considers human aspects inclusively in the design and development process of intelligent solutions. Since, modeling a range of user diversity parameters, e.g. intrinsic human factors, demographics, motivation and self-regulation, and consolidating these in adaptive and personalized systems still remains an open challenge; and viable long-term solutions are yet to be found.

This is especially relevant for systems that promote learning and behavior change which require more holistic human-centered adaptation and personalization. Successful approaches could be realized by a) defining more accurate human-centered models based on intrinsic human factors and abilities, such as perceptual, personality, visual, cognitive, and emotional factors as expressed by the theories of individual differences, as well as on other inherent or more recognizable diversity user characteristics like age, culture, status, motivation, expertise, self-actualization, socio-cultural behavior, etc.; and b) creating intelligent algorithms, interaction principles and smart interfaces that can handle the increasing computational complexity, behavioristic patterns, data structures and the high volume of the generated multi-purpose information.

This article collection is primarily inspired by the International workshop HAAPIE (http://haapie.cs.ucy.ac.cy), held annually in conjunction with the ACM UMAP Conference. It encloses a selection of extended high-quality papers that have been presented in the series of HAAPIE workshops and original unpublished research works that have a considerable contribution and influence in the field. Accordingly, this special issue contains nine contributions discussing interesting ideas in the areas of adaptive information visualization and analytics; human factors and taxonomies; biases and social media; mental/physical health, persuasive technologies and behavior change; e-commerce, motivation and evaluation; learning activity and emotions; and opinion formation on the internet and personality traits. More specifically, Steichen and Fu discuss that cognitive styles have a direct impact when users engage with tasks that include Information Visualizations, and that there are distinctive differences on individual aid choices and preferences, motivating the development of adaptive Information Visualization systems. In the same line of research, Poetzsch et al. argued in their work that data visualizations should be adapted to both the user and the context employing a user model that combines user traits, states, strategies, and actions. They proposed a taxonomy for visualization recommendations paving the way for adaptive data visualizations in analytics. Aïmeur et al. analyzed the motivations and cognitive biases which are frequently exploited by deceptive attackers in Social Network Sites, proposing some countermeasures for each of these biases to provide personalized privacy protection against deceivers. Main concern of Alqahtani et al. was to understand how the persuasive strategies promote mental health. They provide a comprehensive review in the field, and by examining the relationship between mental health apps effectiveness and the persuasive strategies offer design recommendations. In this research direction, Aldenaini et al. provide a systematic review of persuasive technologies for promoting physical activity and reducing sedentary behavior. They answer some fundamental questions in terms of design and effectiveness evaluation, behavioral theories, etc., and reveal the pitfalls and gaps in the present literature. Adaji et al., investigated which factors could tailor persuasive strategies in e-Commerce so to be more effective. They propose the use of shoppers' online shopping motivation in tailoring six commonly used persuasive strategies; showing that persuasive strategies influence e-commerce shoppers differently based on their shopping motivation. Esteller-Cucala et al. discuss five experimentation pitfalls, especially for online controlled experiments (A/B tests), initially identified in an automotive company's website—followed by other sectors, which are highly probable to appear when evaluating personalization features. Alhathli et al. investigates how humans adapt next learning activity selection to learner personality, emotional stability and competence to inspire an adaptive learning activity selection algorithm. The algorithm selects learning activities with varying assumed and taught knowledge adapted to learner characteristics. Lastly, Burbach et al. created an agent-based model and simulated message spreading in social networks to investigate which factors influence whether a user disseminates information or not. Findings reveal that the network type has only a weak influence on the distribution of content, whereas the message type has a clear influence on how many users receive a message.

The accepted manuscripts convey a representative angle of the theoretical dimensions and practical insights when considering individual differences during the process of user modeling, adaptation, and personalization in various research domains and application fields. Their outcomes and suggestions underline the evident potential and capabilities of the related intelligent solution to keep the user *haapie* in the end!

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This Research Topic is in memory of Prof. George Samaras (1959–2018) who passionately believed that human factors played key role and should be taken into account in adaptation and personalization. Not only was George Samaras a key driver behind initiating the HAAPIE workshop series, he led research that provided some of the early examples of user-adaptive systems that take into account human factors such as cognitive ability and age.

Conflict of Interest: The author PG is employed by SAP SE, Walldorf, Germany.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 01 November 2019
doi: 10.3389/frai.2019.00022

Towards Adaptive Information Visualization - A Study of Information Visualization Aids and the Role of User Cognitive Style

Ben Steichen[1]* *and Bo Fu*[2]

[1] Department of Computer Science, California State Polytechnic University, Pomona, CA, United States, [2] Department of Computer Engineering and Computer Science, California State University, Long Beach, CA, United States

Information Visualization systems have traditionally followed a one-size-fits-all model, whereby the same visualization is shown to each user, without taking into consideration an individual user's preferences, abilities, or context. By contrast, given the considerable cognitive effort involved in using Information Visualizations, this paper investigates the effect of an individual user's cognitive style on Information Visualization performance. In addition, this paper studies several interactive "visualization aids" (i.e., interactive overlays that can aid in visualization comprehension), as well as the effect of cognitive style on aid choices and preferences. The results from a user study show that cognitive style plays a significant role when performing tasks with Information Visualizations in general, and that there are clear differences in terms of individual aid choices and preferences. These findings also provide motivation for the development of adaptive and personalized Information Visualization systems that could better assist users according to their individual cognitive style.

Keywords: information visualization, adaptation, cognitive style, interaction, human-centered computing, personalization

OPEN ACCESS

Edited by:
Panagiotis Germanakos,
SAP SE, Germany

Reviewed by:
Lei Shi,
Durham University, United Kingdom
Elvira Popescu,
University of Craiova, Romania

***Correspondence:**
Ben Steichen
bsteichen@cpp.edu

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *13 August 2019*
Accepted: *15 October 2019*
Published: *01 November 2019*

Citation:
Steichen B and Fu B (2019) Towards Adaptive Information Visualization - A Study of Information Visualization Aids and the Role of User Cognitive Style. Front. Artif. Intell. 2:22. doi: 10.3389/frai.2019.00022

INTRODUCTION

One of the most powerful ways to help humans perform cognitive work is to support them with interactive visualizations, particularly through computer-generated Information Visualizations (Spence, 2001; Ware, 2004). Given the unprecedented amount of information now available to people, organizations, and communities, the use of Information Visualization systems has become ubiquitous for diverse populations across a wide variety of activities, such as reading newspaper articles, exploring scientific data, or making business decisions.

Traditionally, Information Visualization systems have followed a one-size-fits-all model, whereby the same (often non-interactive) visualization is shown to each user, without taking into consideration an individual user's preferences, abilities, or context. By contrast, in fields outside of Information Visualization, there are ample established examples of successfully designing systems that are personalized to individual users, such as in Personalized Information Retrieval (Steichen et al., 2012), Adaptive Web systems (Steichen et al., 2012), or Adaptive E-learning (Jameson, 2008).

In the field of Information Visualization, such research regarding interaction, adaptation, and personalization has emerged only recently, showing that individual user characteristics may have an impact on Information Visualization effectiveness, and that there is potential for the development of adaptive and personalized Information Visualization solutions. As with any development of such systems, researchers have focused on (i) determining what specific characteristics may play a role in a user's interaction with a system, and (ii) devising mechanisms to help users.

In this paper, we similarly focus on both of these aspects, and extend prior work by (i) investigating the effect that a user's cognitive style may have on their use of Information Visualizations, and (ii) investigating several general Information Visualization "aids" that may be added to an existing visualization to assist users during typical tasks (i.e., interactive overlays that can aid in visualization comprehension).

The focus on a user's cognitive style is based on several related research works outside of Information Visualization, which have shown that this user characteristic can have significant effects on a user's processing of visual information (Witkin et al., 1975; Mawad et al., 2015; Raptis et al., 2016). Since using Information Visualizations consists of complex cognitive activities that make significant use of visual information, we hypothesize that this characteristic may therefore have a significant impact.

The focus on general visualization "aids" that are added to an existing visualization (i.e., visualization overlays) is motivated by the fact that prior work has so far mostly concentrated on visualization highlighting effects (Carenini et al., 2014) (i.e., highlighting specific data points), which by definition require the system to know exactly which data points the user is most interested in. While this is a valid assumption in the case of systems that, for example, present visualizations along with a textual description (e.g., a newspaper article that is accompanied by a visualization), this is not generally the case. In particular, users may be engaged in several different tasks on a single visualization, and the visualization system developer/provider may not know which aspects or data points the user is focused on at any given time. More general Information Visualization aids, such as grid overlays or added labels, may therefore be more appropriate in such cases.

In order to investigate these Information Visualization aids, as well as the role of a user's cognitive style, this paper presents a user study where participants interacted with two common Information Visualizations, namely bar graphs and line graphs, and five different visualization aids. The specific research questions that this user study aims to answer are:

1. To what extent does a user's cognitive style play a role when performing tasks with Information Visualization systems? (RQ1)
2. In general, which Information Visualization Aids do users choose the most, and which are considered most helpful by users? (RQ2)
3. Does cognitive style play a role in aid choice and subjective usefulness? (RQ3).

RELATED WORK

Research on the effect of, and adaptation to, individual user characteristics has long been established in fields outside of Information Visualization. Prominent examples include Adaptive Hypermedia (Steichen et al., 2012), Personalized Information Retrieval (Steichen et al., 2012), and Adaptive e-Learning (Jameson, 2008). In each of these fields, the first step is to identify an influential user characteristic, followed by research on how to best support each individual user in a personalized manner. For example, the goal of many Personalized Information Retrieval systems is to personally tailor search results to each individual user (Steichen et al., 2012). In order to achieve this goal, systems may employ a range of techniques to, for example, (i) gather individual user interests from prior queries and result selections, in order to (ii) tailor retrieval algorithms to re-rank search results based on these interests. Likewise, Adaptive e-Learning systems may (i) gather a user's knowledge through tests or interaction patterns, in order to (ii) provide a personalized path through the learning material.

Human Factors and Information Visualization

Besides the above examples of "traditional" user characteristics (e.g., user interests or prior knowledge), more recent work has also investigated the effect of human factors, such as cognitive processing capabilities (Germanakos et al., 2009). In particular, one human factor that has been consistently shown to influence human behavior is the high-level cognitive process of cognitive style. Specifically, according to the (FD-I) theory, *Field Dependent* people tend to have difficulties in identifying details in complex scenes, whereas *Field Independent* people easily separate structures from surrounding visual context (Witkin et al., 1975). This characteristic has been shown to have significant effects in several areas outside of Information Visualization, for example when playing games (Raptis et al., 2016) or making purchasing decisions (Mawad et al., 2015). Specifically, gamers have been shown to have varying completion speeds and behavioral patterns depending on this characteristic (Raptis et al., 2016). Likewise, users showed different information processing behaviors when reading product labels (Mawad et al., 2015). Recent research has shown that such differences can even be implied from eye gaze data (Mawad et al., 2015; Raptis et al., 2017). Given the intricate connection of this user characteristic with visual tasks, our paper therefore hypothesizes that it may also have an influence on Information Visualization use.

The effect of individual user differences and human factors on behaviors with Information Visualizations has only been studied very recently. Most notably, there are a number of examples showing that there is an effect of personality, cognitive abilities, and expertise on a user's performance with (and preference for) different visualizations (Velez et al., 2005; Green and Fisher, 2010; Ziemkiewicz et al., 2011; Toker et al., 2012; Carenini et al., 2014; Luo, 2019). For example, results in Ziemkiewicz et al. (2011) showed that users with an internal locus of control performed poorly with Information Visualizations that employ a containment metaphor, while those with an external locus of control showed good performance with such systems. This finding provided motivation for the tailoring/selection of different Information Visualizations for different users, depending on their locus of control. Similarly, results in Toker et al. (2012) showed that user cognitive abilities, such as perceptual speed and working memory had an influence on visualization preferences and task completion time. Most recently, Luo (2019) investigated user cognitive style along the visualizer-verbalizer dimension (Richardson, 1977; Riding, 2001), where individuals were distinguished as either preferring their visual or verbal subsystem. Based on this distinction, results showed that verbalizers preferred table representations of data, whereas visualizers preferred graphical representations (i.e., data visualizations). However, the effects of a user's cognitive style according to the (FD-I) theory have, to the best of our knowledge, not been explored in Information Visualization, despite its proven effect on visual tasks in other fields (Mawad et al., 2015; Raptis et al., 2016, 2017). Our paper addresses this research gap by studying the effect of cognitive style according to the (FD-I) theory.

Interaction, Adaptation, and Personalization

As with the study of the effects of individual user differences, there have been extensive studies of novel interaction and adaptation mechanisms outside of the area of Information Visualization. For example, related work has looked at a variety of adaptation techniques, such as display notifications (Bartram et al., 2003), hint provisions (Muir and Conati, 2012), search result reranking (Steichen et al., 2012), or adaptive navigation (Steichen et al., 2012).

In Information Visualization, the most common interaction and adaptation technique has typically been to recommend alternative visualizations (Grawemeyer, 2006; Gotz and Wen, 2009). More recently, Kong et al. developed a system that could dynamically add overlays to a visualization in order to aid chart understanding (Kong and Agrawala, 2012). In particular, the developed overlays were "reference structures" (e.g., grids), "highlights" (e.g., highlighting a particular bar in a bar graph), "redundant encodings" (e.g., data labels), "summary statistics" (e.g., mean line), and "annotations" (e.g., providing comments on particular data points). However, no studies were performed to investigate the relative benefits, drawbacks, or individual user preferences.

Most closely to our work, Carenini et al. (2014) proposed the personalization of visualizations that a user currently engages with [rather than providing personalized recommendations for alternative visualizations as in Grawemeyer (2006) and Gotz and Wen (2009)]. The actual adaptation techniques proposed in Carenini et al. (2014) were inspired by an analysis of classical Infovis literature (Bertin, 1983; Kosslyn, 1994), as well as a seminal taxonomy on "visual prompts" from Mittal (1997). Similar to the abovementioned "overlay techniques" in Kong and Agrawala (2012), these "visual prompts" were a collection of visualization overlays and parameters that could be added or changed on a visualization, either interactively or adaptively. In particular, Carenini et al. (2014) focused on a subset of "visual prompts" from Mittal (1997) that could be used for highlighting specific data points that are relevant to the user's current task. The chosen techniques in Carenini et al. (2014) therefore require a system to have exact knowledge of the user's task, e.g., knowing exactly which two data points on a graph the user is interested in comparing with each other. This assumption is based on the idea of "*Magazine Style Narrative Visualization*" as presented in Segel and Heer (2010) and Kong et al. (2014), where the visualization is meant to accompany a known textual narrative (Segel and Heer, 2010).

However, this assumption of knowing the exact elements of interest to the user cannot be guaranteed for visualizations in general. By contrast, the work in our paper focuses on visual prompts, called "visualization aids" in our paper, that can be added to a visualization without knowing the exact data points that a user is interested in (e.g., reference structures, such as grids), thereby making them task-independent and applicable for different types of scenarios. In addition, our work explores the effect of cognitive style on aid usage and preferences.

EXPERIMENTAL SETUP

In order to study user behaviors and preferences with regards to different interactive visualization aids, as well as the effects of a user's cognitive style, we conducted a laboratory experiment involving two different visualizations, as well as five different visualization aids. Overall, 40 participants took part in the study, which consisted of a series of visualization tasks to be completed using the given visualizations. The following paragraphs describe the visualizations and aids used in the study, the study tasks and procedure, as well as the participant recruitment and data analysis.

Visualizations and Aids Used in the Study

The study was conducted using two visualization types, namely bar graphs and line graphs. The choice for these visualizations was based on their ubiquitous adoption across different fields and media, as well as some use in prior work on user differences (e.g., bar graphs in Toker et al., 2012).

For each of the visualizations, five visualization aids were available to participants, which were largely based on the "visual prompts" taxonomy presented in Mittal (1997) (and also used in Carenini et al., 2014). In particular, each of these visualization aids fall into the "overlay possible (*ad-hoc*)" category (i.e., aids that can be overlaid dynamically, even by a third-party software as presented in Kong and Agrawala, 2012), as opposed to "planned with original design" (i.e., requiring significant changes to the graph that could only be made if included in advance by the original visualization designers, e.g., axis change, typeface change). As such, they also adhere to the "reference structures" and "redundant encodings" categories from the taxonomy in Kong and Agrawala (2012).

The choice for these particular types of aids was based on the fact that they can be used as an overlay on an existing visualization (and may therefore be used as an interactive or adaptive help for users), and that they do not require any knowledge of the user's focus on any particular data point. In addition, all of the chosen aids were applicable to both bar graphs and line graphs (and potentially other visualizations), thereby also allowing an analysis of any effects of visualization type.

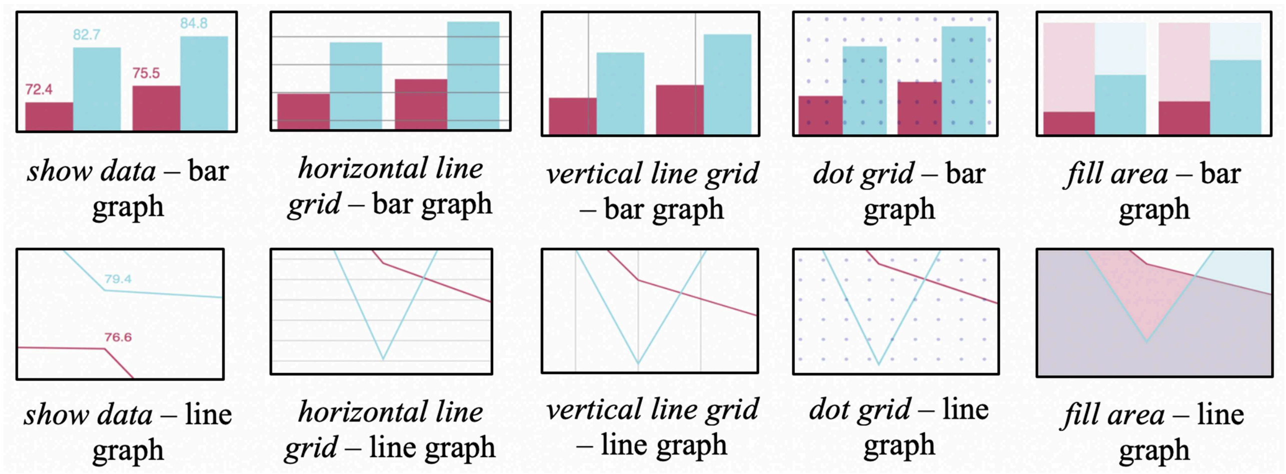


FIGURE 1 | Visualization aids used in the study.

Figure 1 shows all five aids, for both bar graphs and line graphs. Specifically, the aids were:

- *show data*—adding the exact data point values above the respective bar/line. The hypothesis for this aid is that it helps users who have difficulties in comparing two data points using purely graphical representations.
- *horizontal line grid*—overlaying a horizontal grid. The hypothesis for this aid is that it helps users in comparing specific points across a graph through additional structure (e.g., for comparing the height of two bars that may be on opposite sides).
- *vertical line grid*—overlaying a vertical grid. The reason for including this aid in the study is the hypothesis that some participants may like to combine horizontal and vertical lines to form additional structure that may help in dissecting a visualization.
- *dot grid*—overlaying a dot grid. This aid is included as an alternative to the above solid grids, as it may be preferred as a less intrusive option.
- *fill area*—adding a shaded complement in a bar graph/adding a shaded area underneath a line for the line graph. This aid thereby represents an alternative reference structure aid. The hypothesis is that some users may prefer the provided additional visual representations, e.g., some users may always prefer to compare shorter or longer bars, or use the visual cues provided by overlaps in the line graphs.

Each of these aids could be toggled on and off by users through checkboxes. In addition, the system allowed users to toggle multiple aids (i.e., overlay) at any given time. Also, the order of aid checkboxes was randomized on a per-participant basis, to minimize any ordering effects while still maintaining a consistent interface for each individual participant.

Experimental Tasks

Each participant performed a set of tasks related to two standard datasets drawn from Data.gov, namely the Diabetes Data Set[1] and the Los Angeles Crime [2] dataset. A task consisted of a question, a corresponding graph, and a set of possible answers (see **Figure 2**).

Half of the questions required the choice of only one answer (using radio buttons), with the other half allowing the choice of multiple correct answers (using checkboxes). The tasks were designed to be of varying type and complexity. In particular, the questions were based on the taxonomy of task types presented in Amar et al. (2005), and consisted of "Retrieve Value," "Compute Derived Value," "Filter," and "Find Extremum" tasks.

Furthermore, the graphs were either of "Low Information Density," which showed only two series (as in **Figure 1**), or "High Information Density," which showed seven series (see **Figure 3** for an example of a "High Information Density Bar Graph"). This distinction was included to facilitate the analysis of potential effects of information density on aid usage. For example, it may be the case that aids are not considered important for "Low Information Density" graphs, while some/all participants may see a benefit of aids for "High Information Density" graphs.

Procedure

Each participant followed the same study procedure, which started with the agreement to a consent form. This was followed by demographic questionnaires regarding participant age, gender, as well as self-reported experience/expertise with different types of visualizations. Specifically, they were asked how often they work with high/low information density bar/line graphs, on a scale from 1 (never) to 5 (very frequently).

Each participant was then presented with the same two practice tasks (one per visualization type, each using high information density), where they were encouraged to familiarize

[1] https://catalog.data.gov/dataset/diabetes
[2] https://catalog.data.gov/dataset/crime-data-from-2010-to-present-c7a76

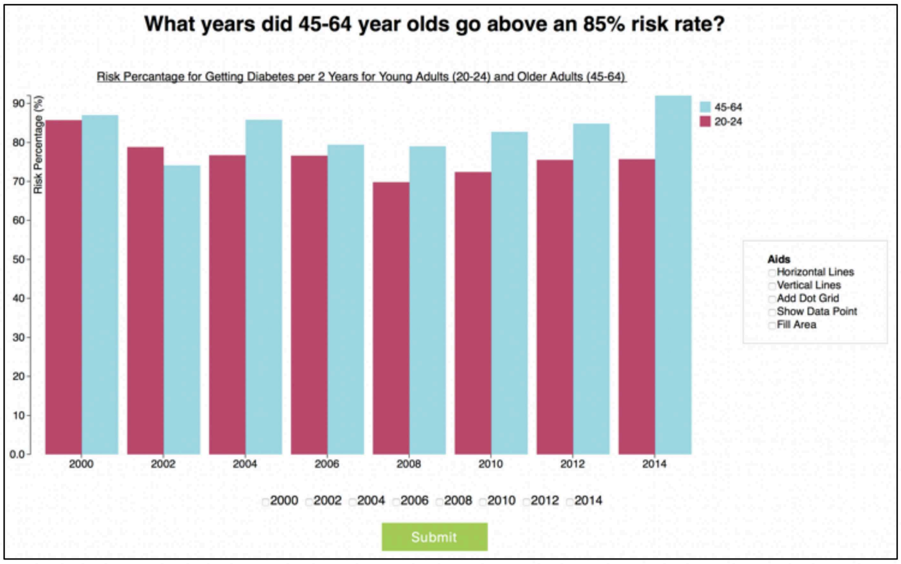


FIGURE 2 | Sample task from the diabetes data set, with low information density bar graph.

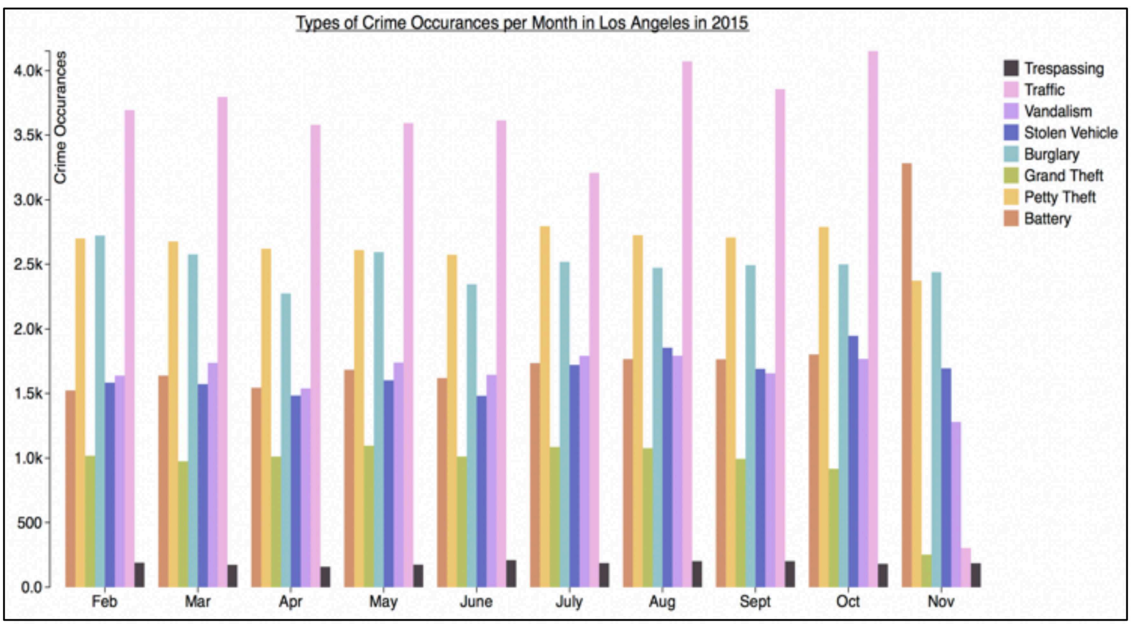


FIGURE 3 | Sample high information density graph for the Los Angeles crime data set.

themselves with the graph layouts and question/answer types, as well as to try out all of the aids.

Following the practice tasks, participants performed 50 tasks (25 with each visualization; total of 20 high information density, 30 low information density), where graph type, task question, and information density were all counterbalanced across participants to avoid any ordering effects. For each task, the participant's time was recorded, along with all mouse clicks.

After all tasks were completed, participants filled out a post-task questionnaire, where they noted their perceived usefulness of the different aids (on a 5-point Likert scale).

Lastly, users' cognitive styles according to the FD-I theory were measured through the Group Embedded Figures Test (GEFT)[3] (Oltman and Witkin, 1971), which is a reliable and

[3]For this study, we used the online version of the test—https://www.mindgarden.com/105-group-embedded-figures-test-a-measure-of-cognitive-style

validated test that has been frequently used in prior research (e.g., Mawad et al., 2015; Raptis et al., 2016).

The average session lasted ~1 h, and each participant was compensated with a $20 gift voucher.

Participant Recruitment and Demographics

40 participants were recruited by the authors through University mailing lists. The age range was between 18 and 77 (average of 28 years), 24 participants were female, and 16 were male. The participants consisted of students, faculty, and administrators. There was a balanced distribution across colleges and departments (e.g., arts, business, engineering, science), thereby ensuring minimized bias toward any domain-specific population. The average GEFT score was 13.75/18 (SD = 4.24), suggesting the population was slightly biased toward field independence. The average self-rated expertise of participants was 3.18 (SD = 0.93) out of 5 for "Simple Bar" visualizations, 2.50 (SD = 1.04) for "Complex Bar" visualizations, 3.40 (SD = 0.87) for Simple Line visualizations, and 2.80 (SD = 0.88) for Complex Line visualizations.

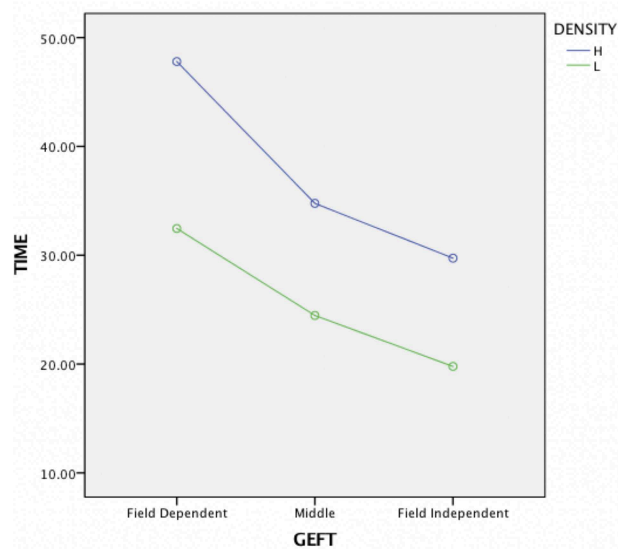


FIGURE 4 | Effects of cognitive style (GEFT) and Information density on task time.

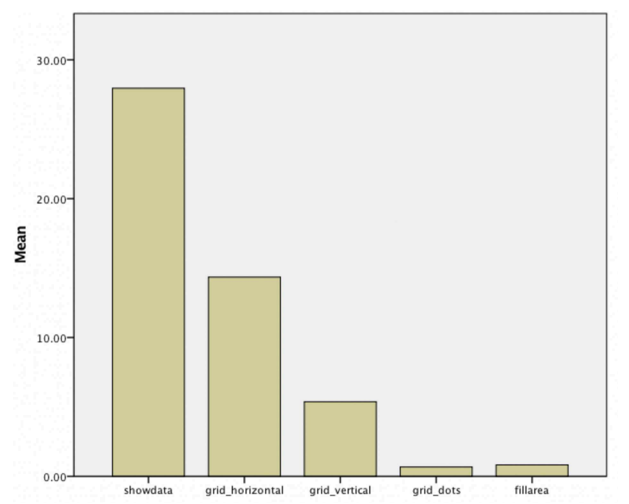


FIGURE 5 | Average use of aids per participant (across 50 tasks).

Data Analysis

All data was analyzed using General Linear Models (GLM), which are a generalization of ordinary linear regression models (i.e., a generalization that incorporates a number of different statistical models, such as ANOVA, ANCOVA, MANOVA, MANCOVA, ordinary linear regression, *t*-test, and *F*-test) (Field, 2009). The independent measures used in the models were graph type, information density, user cognitive style, and user expertise. The dependent measures were accuracy (whether participants submitted the correct answer), task time (measured from the start of a task to pressing the submit button), aid count (how frequently participants made use of specific aids), and subjective preferences (from the post-task questionnaire).

RESULTS

This section presents the general results for each of the dependent measures, i.e., accuracy, time, aid count, and preferences. In addition, this section reports on our analysis of the influence of a user's cognitive style on these measures. Expertise (as measured through the self-reported questionnaire) did not have an effect on any of the measures and is therefore not reported further.

Accuracy

Overall, the mean accuracy across all participants was very high at 87% (43.72 correct tasks out of 50). It therefore appears that participants may have been taking as much time as needed to get the correct answer, i.e., they may have been penalizing time for accuracy (similar to results found in Toker et al. (2012) and Carenini et al. (2014). No effects were found for any of the independent factors on this measure, most likely because of the overall high accuracy (and therefore lack of variance) across participants.

Time

Participants took on average 29.75 s to complete a task, with a standard deviation of 19.47 s. As expected, high information density tasks took considerably longer (34.76 s) compared to low information density tasks (23.56), and this difference was statistically significant ($F_{1,39} = 44.31$, $p < 0.001$). Likewise, graph type played a small role, with participants taking slightly longer with Bar graphs (30.30 s) compared to Line graphs (28.03). This difference was also statistically significant ($F_{1,39} = 4.07$, $p < 0.05$). In addition, there was a statistically significant ($F_{1,39} = 10.583$, $p < 0.05$) interaction effect between graph type and information density, with high information density tasks showing a difference between the two graphs, while both graphs performed almost equally on low density tasks.

A user's cognitive style, as measured through the GEFT, had a statistically significant effect on a user's time on task. Specifically, participants with high field independence scores had statistically significantly faster times than participants with low field independence scores ($F_{1,39} = 187.60$, $p < 0.001$). When using a three-way split [as recommended by Cureton (1957)], which differentiates between Field Independent (FI-upper 27%), Field Dependent (FD-lower 27%), and Middle participants, FD participants ($N = 10$) were found to take 36.5 s, Middle participants ($N = 19$) 27 s, and FI participants ($N = 11$) 23.5 s (see **Figure 4**). This finding was slightly more pronounced for high density tasks compared to low density tasks ($F_{1,39} = 6.70$, $p < 0.01$). Lastly, aid use did not lead to any statistically significant performance increases for FD or FI participants.

Aid Count

Overall, participants turned on an aid 1,967 times (average of 49.17 per participant). The most popular aids were *show data* (27.95 uses on average) and *horizontal line grid* (14.35), while the other aids were less popular, with 5.3 uses for *vertical grid line* and only 0.67 and 0.82 uses for *dot grid* and *fill area*, respectively (see **Figure 5**). This difference between aids was statistically significant ($F_{4,39} = 32.22$, $p < 0.001$).

This general trend was found across both graphs. Additionally, there was a statistically significant effect of graph type on the use of the *vertical line grid* ($F_{1,39} = 5.23$, $p < 0.03$), with this aid being more popular for the line graph (average of 5.37 uses per participant) compared to the bar graph (average of 1.17 uses per participant). Task Density, however, showed no statistically significant effect overall on any aid counts.

A participant's cognitive style had a statistically significant effect on aid count, with FD participants making substantially more use of aids compared to FI participants ($F_{1,39} = 7.25$, $p < 0.01$). Specifically, when using a three-way split, FD participants were found to use 63.5 aids on average, Middle participants used 46 aids, and FI participants used 45.36 aids (see **Figure 6**). When further breaking down these results, it was found that the difference between FD and Middle/FI participants was particularly striking for the *show data* aid. Specifically, FD participants used *show data* aids 43.5 times, vs. only 22.36 times for middle and 23.45 times for FI participants (see **Figure 7**). This result was found to be statistically significant ($F_{1,39} = 6.81$, $p < 0.003$). Aid use for the other aids was almost equal between FD and FI participants, and, in fact, FI users chose *horizontal grid* slightly more often than FD users (14.9 vs. 13.9, ns). Task Density only had a marginally significant interaction effect with cognitive style ($F_{1,39} = 1.85$, $p < 0.055$), with FD participants having a slightly more elevated use of aids during high density tasks.

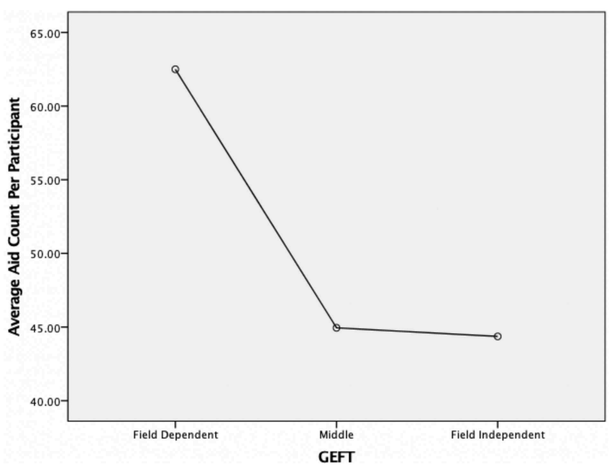


FIGURE 6 | Effect of cognitive style on aid count—overall.

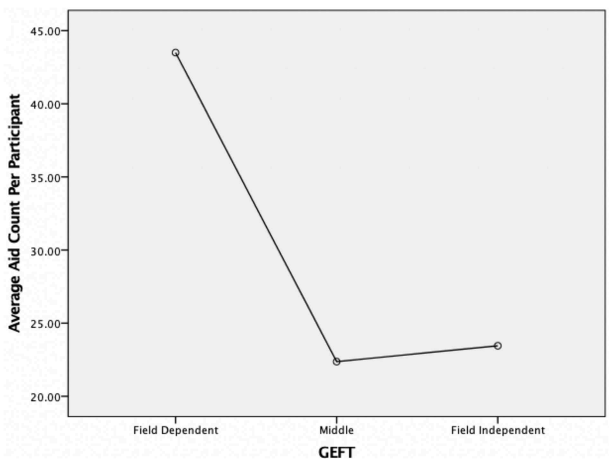


FIGURE 7 | Effect of cognitive style on aid count—*show data* aid only.

Preferences

The analysis of participants' subjective preferences for the different aids (from the post-task questionnaire) revealed similar results to the aid count analysis above. In particular, both *show data* and *horizontal line grid* were considered the most useful, with usefulness scores of 4.35 and 3.93, respectively (5 being very useful, and 1 not being useful at all), while none of the other aids (or *no aid*) were considered useful (2.4 for *vertical line grid*, 1.9 for *dot grid*, 1.7 for *fill area*, and 2.0 for *no aid*) (see **Figure 8**). This difference between aids was again statistically significant ($F_{5,39} = 77.36$, $p < 0.001$).

Likewise, there was a statistically significant effect of graph type for the *vertical line grid* ($F_{1,39} = 100.56$, $p < 0.001$), with this aid being considered more useful for the line graph (3.26) compared to the bar graph (1.54).

As with aid count, there was a statistically significant effect of cognitive style on perceived usefulness, with FD participants reporting greater usefulness across all aids overall. Again, *show data* in particular showed a statistically significant effect, with a rating of 4.6 for FD participants, 4.35 for Middle, and 4.1 for FI participants ($F_{1,39} = 4.77$, $p < 0.03$).

SUMMARY AND DISCUSSION

The study has revealed a number of interesting findings regarding the three research questions posed in the introduction. This

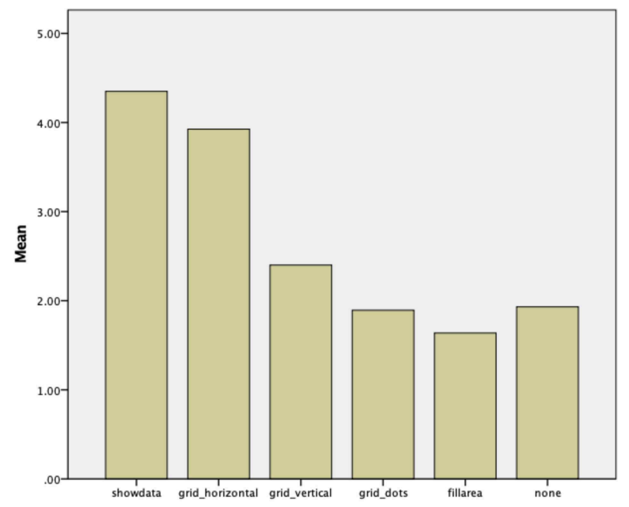


FIGURE 8 | Perceived aid usefulness (5 = very useful, 1 = not useful at all), "none" refers to the usefulness of not having any aids overlaid.

section first provides a summary of these findings, as well as implications for design and potential adaptation, followed by a discussion of the limitations of the study.

Firstly, the results from the study have shown that cognitive style indeed plays a significant role when performing tasks with Information Visualizations (RQ1). In particular, FI users have been shown to be significantly faster at completing tasks compared to FD users. This is in line with similar work in Information Visualization regarding other cognitive measures, such as perceptual speed or working memory (e.g., Velez et al., 2005; Toker et al., 2012; Carenini et al., 2014). Similar to design suggestions in such prior work, this may indicate that FD participants may be in particular need to receive additional help in performing visualization tasks, such as through adaptive or personalized aid additions or recommendations. In contrast to task time, however, the study did not reveal any differences in accuracy, possibly due to participants sacrificing task time for accuracy (again, in line with prior work on other user characteristics (e.g., Toker et al., 2012; Carenini et al., 2014).

In general, there were also clear differences in terms of visualization aid choices and preferences (RQ2), with *show data* and *horizontal grid lines* being most used and subjectively preferred. When designing interactive visualizations with visualization aids, it may therefore be advisable to use one of these two options (or both). The fact that *vertical grid lines* were less used and preferred is understandable (especially for the bar graph), given that they provide less support in judging data point values (since they only produce additional partitioning of the graph). As previously mentioned, the reason for including this in the study was the hypothesis that some participants may like to combine horizontal and vertical lines to form additional structure that may help in dissecting a visualization. However, this seems not to have been the case. The low use of *fill area* or *dot grid* is less clear, although it is conceivable that, at least for the *fill area* aid, high information density tasks may have suffered too much of a performance decrease (due to the many overlaps in line graphs, and the high number of different colors/shades for bar graphs). However, *post-hoc* analyses did not find an effect for information density on this particular aid. The *dot grid* area had been included in the study as an alternative to the solid grids, as it may have been preferred as a less intrusive option. However, it appears that it was not judged to be useful. This suggests that, if grids are to be added, they should be solid lines in order to provide better support for users.

Cognitive style was also shown to play an important role in visualization aid choice and subjective usefulness (RQ3), with FD users making significantly more use of visualization aids during their tasks. Likewise, they clearly noted them more as being useful in completing the tasks, as shown through the final survey. While our study was not able to detect a performance increase when using aids, the fact that participants continued to choose them throughout the study suggests that they felt a benefit, even if just in terms of subjective experience. This suggests that there may be implications of cognitive style to information visualization design, with FD users potentially benefitting most from a system that perhaps adds these aids by default, or one that adds them adaptively (or provides them as recommendations). In terms of specific aid choices, it was shown that FD participants made significant use of the *show data* aid (i.e., the aid that overlays the actual data values), and that they strongly considered this aid to be useful. This suggests that FD participants may have more difficulties using purely visual representations of data, and that they prefer to have additional numerical data displayed in the visualization. This is in line with the general definitions of cognitive style along the FD-I dimension, i.e., FD participants struggling to identify details in complex visual scenes. *Grid horizontal* was also chosen by FD participants to a certain degree, but the fact that it was used significantly less suggests that the added structure through additional visual objects was not appreciated as much by such users. As with research on other forms of cognitive style (e.g., along the verbalizer-visualizer dimension), this may again suggest that additional (non-visual) forms of cognitive aids should be explored for FD users. FI users chose aids significantly less often overall (particularly the *show data* aid), which may potentially suggest that such users might prefer the option to interactively turn on aids themselves, rather than systems where aids are turned on by default. However, this hypothesis requires further research (discussed below).

Lastly, there are a number of limitations of the study, some of which may be addressed in future work (also discussed in the following section). First of all, the study consisted of a laboratory study with 40 participants, 2 visualizations, and 5 visualization aids. While the number of participants was in line with similar prior work (e.g., Bartram et al., 2003; Velez et al., 2005; Grawemeyer, 2006; Green and Fisher, 2010; Toker et al., 2012; Carenini et al., 2014; Raptis et al., 2016) and provided sufficient strength to reach statistical significance for several effects, a larger participant pool may have enabled the discovery of additional effects, as well as the study of a larger set of visualization types and aids. Specifically, there are many other basic visualizations and aids that could be studied, including the various types of visualizations and aids proposed in Kong and Agrawala (2012). While the purpose of our study was to focus on particularly common visualizations (bar and line graphs) and aids (e.g., grids, labels, etc.), such future investigations of additional basic visualizations could thereby potentially identify which specific (aspects of) visualizations would most benefit from providing aids to users. Likewise, there are many more complex and/or domain-specific visualizations, aids, and tasks that could be studied (e.g., specific visualizations and tasks for decision-making, such as in Conati et al., 2014). While the focus of our study was on basic, common, and domain-independent visualizations and tasks, more complex visualizations, aids, and tasks could potentially bring out even stronger results. Specifically, complex visualizations may elicit more aid usage from users, and also potentially reveal even bigger differences depending on cognitive style.

CONCLUSIONS AND FUTURE WORK

Overall, the results from the study provide valuable information regarding the role of user cognitive style in Information Visualization, as well as initial implications for the design of Information Visualization aids. In particular, it was shown that cognitive style has a significant impact on visualization task performance, that different visualization aids are chosen to different degrees, and that cognitive style has a significant influence on which aids are chosen and considered most helpful.

The paper thereby provides further motivation for the development of adaptive and personalized Information Visualization systems, as previously proposed in related work (e.g., Grawemeyer, 2006; Gotz and Wen, 2009; Toker et al., 2012; Carenini et al., 2014; Steichen et al., 2014). In particular, the paper provides the first results that motivate the adaptation and personalization of Information Visualization systems depending on a user's cognitive style, through the use of visualization aids.

Based on this motivation, there are several avenues for further research. As discussed in the previous section, there are many additional visualizations and aids that may be studied, which may uncover whether there are particular (aspects of) visualizations that elicit differences depending on cognitive style. Such studies of larger sets of visualizations would require the recruitment of larger participant pools, in order to ensure the same statistical power. In addition, the study of more complex and/or domain-specific visualizations and tasks may provide further insights into the role of cognitive style on different (types of) visualizations, and/or potentially reveal specific application scenarios where aids may be particularly useful.

Furthermore, additional research needs to be conducted in order to study the effects of adding visual aids by default, or adding them adaptively while a user is performing a task (e.g., as in Carenini et al., 2014). While our study did not find a significant performance improvement from aid usage (potentially due to participants needing to spend time to choose and turn on aids), such research may also be able to better quantify such effects. These findings would complement the aid choice and perceived usefulness results from this paper. Since *show data* and *horizontal grid* were by far the most popular aids in our current study, an initial investigation of default/adaptive aid addition may specifically focus on these two visualizations (as well as perhaps additional non-visual aids for FD users).

Furthermore, in order to develop a personalized system, an adaptive aid component would also need to be integrated with a system that can automatically recognize a user's cognitive style, for example using eye gaze data (as in Raptis et al., 2017). Our future research will involve the development of such systems, by extending work in Raptis et al. (2017) to the field of Information Visualization. Given the successful results in Raptis et al. (2017), as well as successful predictions (using eye gaze) of other user types of user characteristics during visualization tasks (e.g., perceptual speed and working memory in Steichen et al., 2014), we hypothesize that this is an achievable task.

Finally, once all components have been developed and evaluated, the final stage of research will involve an investigation of integrated systems that adaptively add or suggest aids based on a user's predicted cognitive style.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Subjects Committee of the primary author's institution. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

REFERENCES

Amar, R. A., Eagan, J., and Stasko, J. T. (2005). "Low-level components of analytic activity in information visualization," in *16th IEEE Information Visualization Conference*, 15–21.

Bartram, L., Ware, C., and Calvert, T. (2003). Moticons: detection, distraction and task. *Int. J. Hum. Comput. Stud.* 58, 515–545. doi: 10.1016/S1071-5819(03)00021-1

Bertin, J. (1983). *Semiology of Graphics*. University of Wisconsin Press.

Carenini, G., Conati, C., Hoque, E., Steichen, B., Toker, D., and Enns, J. T. (2014). "Highlighting interventions and user differences: informing adaptive information visualization support," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1835–1844.

Conati, C., Carenini, G., Hoque, E., Steichen, B., and Toker, D. (2014). Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. *Comput. Graph. Forum* 33, 371–380. doi: 10.1111/cgf.12393

Cureton, E. E. (1957). The upper and lower twenty-seven per cent rule. *Psychometrika* 22, 293–296. doi: 10.1007/BF02289130

Field, A. (2009). *Discovering Statistics Using SPSS. 3rd Edn.* London: Sage Publications.

Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., Belk, M., and Samaras, G. (2009). Towards an adaptive and personalized web interaction using human factors. *Adv. Seman. Media Adapt. Personal.* 2, 247–282. doi: 10.1201/9781420076653-c12

Gotz, D., and Wen, Z. (2009). "Behavior driven visualization recommendation," in *Proceedings of the 14th International Conference on Intelligent User Interfaces, IUI 2009* (Sanibel Island, FL), 315–324.

Grawemeyer, B. (2006). "Evaluation of ERST – an external representation selection tutor," in *Proceedings of the International Conference on Diagrammatic Representation and Inference*, 154–167.

Green, T. M., and Fisher, B. (2010). "Towards the personal equation of interaction: the impact of personality factors on visual analytics interface interaction," in *IEEE Visual Analytics Science and Technology, VAST*.

Jameson, A. (2008). "Adaptive interfaces and agents," in *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. 2nd Edn.* eds A. Sears and J. A. Jacko (CRC Press), 433–458.

Kong, N., and Agrawala, M. (2012). Graphical overlays: using layered elements to aid chart reading. *IEEE Trans. Vis. Comput. Graph.* 18, 2631–2638. doi: 10.1109/TVCG.2012.229

Kong, N., Hearst, M. A., and Agrawala, M. (2014). "Extracting references between text and charts via crowdsourcing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, 31–40.

Kosslyn, S. M. (1994). *Elements of Graph Design*. W. H. Freeman and Company.

Luo, W. (2019). User choice of interactive data visualization format: the effects of cognitive style and spatial ability. *Decis. Supp. Syst.* 122:113061. doi: 10.1016/j.dss.2019.05.001

Mawad, F., Trías, M., Giménez, A., Maiche, A., and Ares, G. (2015). Influence of cognitive style on information processing and selection of yogurt labels: Insights from an eye-tracking study. *Food Res. Int.* 74, 1–9. doi: 10.1016/j.foodres.2015.04.023

Mittal, V. O. (1997). "Visual prompts and graphical design: a framework for exploring the design space of 2-D charts and graphs," in *Proceedings of AAAI/IAAI*, 57–63.

Muir, M., and Conati, C. (2012). "An analysis of attention to student – adaptive hints in an educational game," in *Proceedings of Intelligent Tutoring Systems*, 112–122.

Oltman, P. K, Raskin, E., and Witkin, H. A. (1971). *Group Embedded Figures Test*. Consulting Psychologists Press.

Raptis, G. E., Fidas, C. A., and Avouris, N. M. (2016). "Do field dependence-independence differences of game players affect performance and behaviour in cultural heritage games?" in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play–CHI PLAY'16*, 38–43.

Raptis, G. E., Katsini, C., Belk, M., Fidas, C., Samaras, G., and Avouris, N. (2017). "Using eye gaze data and visual activities to infer human cognitive styles: method and feasibility studies," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP '17)*, 164–173.

Richardson, A. (1977). Verbalizer-visualizer: a cognitive style dimension. *J. Ment. Imagery* 1, 109–125.

Riding, R. J. (2001). The nature and effects of cognitive style. *Perspect. Think. Learn. Cogn. Styles* 47:72. doi: 10.4324/9781410605986-3

Segel, E., and Heer, J. (2010). Narrative visualization: telling stories with data. *IEEE Trans. Vis. Comput. Graph.* 16, 1139–1148. doi: 10.1109/TVCG.2010.179

Spence, R. (2001). *Information Visualisation*. Reading, MA: Addison-Wesley.

Steichen, B., Ashman, H., and Wade, V. (2012). A comparative survey of personalised information retrieval and adaptive hypermedia techniques. *Inform. Process. Manage.* 48, 698–724. doi: 10.1016/j.ipm.2011.12.004

Steichen, B., Carenini, G., and Conati, C. (2014). Inferring visualization task properties, user performance, and user cognitive abilities from eye gaze data. *ACM Trans. Interact. Intell. Syst.* 4:11. doi: 10.1145/2633043

Toker, D., Conati, C., Carenini, G., and Haraty, M. (2012). "Towards adaptive information visualization: on the influence of user characteristics," in *UMAP 2012*, 274–285.

Velez, M. C., Silver, D., and remaine, M. (2005). "Understanding visualization through spatial ability differences," in *IEEE Visualization, VIS 2005*, 511–518.

Ware, C. (2004). *Information Visualization: Perception for Design. 2nd Edn.* San Francisco, CA: Morgan Kaufmann.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. (1975). Field-dependent and field-independent cognitive styles and their educational implications. *ETS Res. Bull. Series* 2, 1–64. doi: 10.1002/j.2333-8504.1975.tb01065.x

Ziemkiewicz, C., Jordan Crouser, R., Yauilla, A. R., Su, S. L., Ribarsky, W., and Chang, R. (2011). "How locus of control influences compatibility with visualization style," in *Proceedings of IEEE VAST 2011*.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 29 November 2019
doi: 10.3389/frai.2019.00026

Manipulation and Malicious Personalization: Exploring the Self-Disclosure Biases Exploited by Deceptive Attackers on Social Media

Esma Aïmeur[1], Nicolás Díaz Ferreyra[2] and Hicham Hage[3]*

[1] Department of Computer Science and Operations Research (DIRO), University of Montreal, Montreal, QC, Canada, [2] Research Training Group "User-Centred Social Media" University of Duisburg-Essen, Duisburg, Germany, [3] Computer Science Department, Notre Dame University-Louaize, Zouk Mosbeh, Lebanon

In the real world, the disclosure of private information to others often occurs after a trustworthy relationship has been established. Conversely, users of Social Network Sites (SNSs) like Facebook or Instagram often disclose large amounts of personal information prematurely to individuals which are not necessarily trustworthy. Such a low privacy-preserving behavior is often exploited by deceptive attackers with harmful intentions. Basically, deceivers approach their victims in online communities using incentives that motivate them to share their private information, and ultimately, their credentials. Since motivations, such as financial or social gain vary from individual to individual, deceivers must wisely choose their incentive strategy to mislead the users. Consequently, attacks are crafted to each victim based on their particular information-sharing motivations. This work analyses, through an online survey, those motivations and cognitive biases which are frequently exploited by deceptive attackers in SNSs. We propose thereafter some countermeasures for each of these biases to provide personalized privacy protection against deceivers.

Keywords: adaptive privacy, awareness, malicious personalization, self-disclosure, cognitive biases, deception, social media

OPEN ACCESS

Edited by:
Panagiotis Germanakos, SAP SE, Germany

Reviewed by:
Stefanos Gritzalis, University of Piraeus, Greece
Silvia Margarita Baldiris Navarro, Universidad Internacional De La Rioja, Spain

***Correspondence:**
Esma Aïmeur
aimeur@iro.umontreal.ca

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *09 September 2019*
Accepted: *11 November 2019*
Published: *29 November 2019*

Citation:
Aïmeur E, Díaz Ferreyra N and Hage H (2019) Manipulation and Malicious Personalization: Exploring the Self-Disclosure Biases Exploited by Deceptive Attackers on Social Media. Front. Artif. Intell. 2:26. doi: 10.3389/frai.2019.00026

1. INTRODUCTION

Nowadays, Social Network Sites (SNSs) like Facebook, Instagram, or Snapchat are widely used for connecting with friends, acquaintances, or even meeting new people. Basically, these sites have become regular meeting places and redefined, to a large extent, the way people create and maintain social relationships (Joinson, 2008; Penni, 2017). Mainly, SNSs allow people to interact simultaneously with a vast network of users and, thereby, maximize their "social capital." Like in the real world, social links in SNSs are reinforced by disclosing more personal information to others. However, the volume and type of content shared online is larger and more diverse than the one revealed offline (Stutzman et al., 2011; Such and Criado, 2018). Moreover, the time people spend sharing information in SNSs has exponentially increased over the last years (Smith and Anderson, 2018). In consequence, SNSs are appealing to individuals with harmful intentions who see these virtual spaces as valuable sources of private information.

In SNSs, privacy as a human practice acquires a high importance since these are spaces in which users make their private life public. That is, users voluntarily disclose their private information to wide and—sometimes untrusted—audiences through the different communication channels available in these platforms (e.g., instant messaging, posts, stories) (Acquisti and Gross, 2006; Boyd, 2010). However, although users in general have reported high concerns about their privacy, they tend to disclose personal information without foreseeing the potential negative effects. Moreover, they often relay on lax privacy settings and consider their online peers as trusted, which increases significantly the chances of being victims of a malicious user. Consequently, users often regret having shared their personal information in SNSs after they suffer unwanted incidents like *cyber-bullying*, *reputation damage*, or *identity theft* (Wang et al., 2011).

Currently, cyber-attacks tend to focus more on human vulnerabilities instead of flaws in software or hardware (Krombholz et al., 2015). For instance, about 3% of Malware attacks exploit technical lapses while the other 97% target the users through social engineering[1]. In order to gain trust and manipulate their victims, social engineers often employ *online deception* as their attack vector (Tsikerdekis and Zeadally, 2014; Krombholz et al., 2015). Particularly, *deceivers* hide their harmful intentions and mislead other users to reveal their credentials (i.e., accounts and passwords) or perform hazardous actions (e.g., install Malware) (Aïmeur and Sahnoune, 2019). For instance, they often impersonate trustworthy entities using fake SNSs accounts to instigate other users on accessing insecure web links and install malicious software. For this, deceivers exploit users' *motivations*, such as financial or moral gain, and employ different *incentive strategies* to mislead them, accordingly (Albladi and Weir, 2016). Such strategies can take the form of a fake link to a cash prize, or a fake survey on behalf of a prominent non-profit organization.

Understanding the users' motivations is fundamental for the design and success of incentive mechanisms. Particularly, motivations have been widely studied and leveraged to increase users' participation in social applications like discussion forums or web blogs (Vassileva, 2012). As a result, several guidelines and patterns have been elaborated on how to design social interfaces that can attract and sustain active contributions in these virtual communities. However, similar principles can be employed in the design of deceptive strategies that mislead users to reveal personal information. Moreover, as in social applications, these incentives can be personalized to each user (victim) to maximize their effect (damage). This process, in which deceivers use the motivations and cognitive biases of their victims to craft their attacks, can be considered as a case of *malicious personalization* (Conti and Sobiesk, 2009).

This work investigates those motivations and cognitive biases that can be exploited for malicious personalization in SNSs. Particularly, it examines which are the self-disclosure motivations and biases that can be leveraged by deceivers to mislead users into revealing private information. Furthermore, this paper analyses (i) which are the incentive strategies used by deceivers in their attacks, and (ii) the link between self-disclosure motivations and specific categories of personal information. To better understand the role that self-disclosure biases (i.e., cognitive and motivational) have in deceptive attacks, we conducted an online survey with 349 participants via Amazon Mechanical Turk (Mturk). Based on our findings, we elaborate on countermeasures oriented to provide personalized privacy protection against deceivers. In particular, we underline how the findings of this work contribute to the development of personalized risk awareness mechanisms.

The rest of the paper is organized as follows. In the next section, related work on online deception is discussed and analyzed. Following, section 3 introduces the theoretical foundations of this paper. Particularly, the use of motivations and incentives for the design of persuasive technologies is discussed together with role of self-disclosure biases in malicious personalization. Sections 4 and 5 elaborate on the design of our online survey and its results, respectively. Next, in section 6, deception countermeasures based on adaptive risk awareness are elaborated, and the limitations of our approach are discussed. Finally, in section 7, we outline the conclusions of this paper and introduce directions for future work.

2. RELATED WORK

Analysing and understanding the logic behind cyber-attacks is fundamental for developing security and data protection countermeasures. Unlike attacks that focus solely on technical vulnerabilities, social engineering attacks target users with access to critical information. That is, they mislead people into disclosing confidential information or even carrying out hazardous actions through influence and persuasion. There are several types of social engineering attacks each of them relying on different technical, physical and social assumptions. Krombholz et al. (2015) analyzed closely a number of well-known and advanced social engineering attacks like phishing, waterholing and baiting, to determine which are their respective underlying assumptions. As a result, they introduced a taxonomy which classifies these attacks according to (i) the communication channel they exploit (e.g., e-mail, cloud, website), (ii) the operator of the attack (i.e., a human or software), and (iii) the strategy they use to approach the victim (i.e., physical, technical or socio-technical). In line with this approach, Aïmeur et al. (2018) introduced a taxonomy which classifies deceptive attacks in SNSs according to their strategy (i.e., *information harvesting*, *social influence*, or *identity deception*). Such a taxonomy also prescribes a set of preventative strategies for each attack category based on state-of-the-art technologies.

As mentioned in section 1, online deception occurs when social engineers employ manipulation and persuasion techniques to mislead their victims. Hence, the success of a deceptive attack will depend, to a certain extent, on the victim's attitude toward manipulation, their risky behavior and their trust in the perpetrator. Such factors were analyzed by Aïmeur and Sahnoune (2019) in the context of online relationships through a survey-based experiment. Among other findings, the study revealed that users who have been involved in an online relationship are more likely to give away their private information when asked for it. Further research has focused on methods for detecting

[1] Estimates of the number of Social Engineering based cyber-attacks into private or government organizations—https://bit.ly/2k5VKmP (accessed 07/09/2019).

fake identities in SNSs (Alowibdi et al., 2015; van der Walt and Eloff, 2017). Particularly, on using behavioral indicators (e.g., absence of profile picture or suspicious online activity) to identify those accounts that may be administrated by deceivers. However, to the best of our knowledge, not much effort has been made on understanding the self-disclosure biases that are exploited by deceivers to craft their attacks. Consequently, this work investigates the effect of these biases under various deceptive scenarios. Particularly, we analyse the role of incentives and motivations when people self-disclose as the value they assign to particular pieces of private information.

3. THEORETICAL BACKGROUND

Following, the theoretical foundations of this work are introduced. Particularly, we discuss the most relevant perspectives on motivation that exist in the literature and their role in the context of deceptive attacks. In line with this, we examine the different self-disclosure motivations and incentive mechanisms that can be leveraged for the elaboration of such attacks. The concepts introduced in this section set the basis for the elaboration of our online survey.

3.1. Motivations and Incentives

Understanding the motivations behind human behavior has guided, to a large extent, the research agenda of disciplines like economics and psychology (Kraut and Resnick, 2012). Each of these disciplines address the issue of motivation under different assumptions related to the rationality of peoples' decisions and the environment in which such decisions are taken. For instance, classical economics considers people as rational agents that interact in an environment in which certain behavior has associated a particular pay-off (positive or negative) (Vassileva, 2012). In this case, incentive mechanisms are designed to ensure that the overall community fulfills a particular goal (e.g., optimizing the joint welfare of all the individuals) without taking into account the diversity of motivations among its members. Hence, this approach emphasizes the benefit of the community as a whole rather than the one of its members.

Behavioral economics, on the other hand, considers people as irrational and investigates the social, cognitive and emotional factors that may influence their actions. Particularly, this approach has shown that many classical mechanisms are not psychologically valid, and therefore fail on explaining the reasons behind peoples' actions, willingness, and goals (Ariely, 2008; Vassileva, 2012). Furthermore, contributions in the area of behavioral economics have nourished principles of user engagement in the design of information systems. One of the most prominent ones is the incorporation of "gamification" elements (e.g., motivational patterns, rules and feedback loops) in social computing applications to increase users' participation (Hamari and Koivisto, 2013). The use of gamification elements is often grounded in psychological theories, such as the *reinforcement theory* (Skinner, 1969) and the *expectancy theory* (Vroom, 1964), which emphasize the influence of external rewards on people's behavior.

Although gamification has been widely explored in the design of social computing applications, it is often questioned because it relies solely on the use of rewards to generate a motivational effect on users. That is, it often overlooks the effect that intrinsic motivations like enjoyment or personal values may have in peoples' behavior (Vassileva, 2012). Moreover, it also neglects the relevance of motivational factors coming from peoples' social environment, such as status and recognition. Consequently, a considerable amount of research focus on developing motivational strategies that elaborate on such intrinsic and social factors (Ling et al., 2005; Burke et al., 2009; Kraut and Resnick, 2012; Chang et al., 2016). Furthermore, approaches on the personalization of incentives have also been introduced to increase users' participation and engagement in social applications (Berkovsky et al., 2012). The main premise of personalized incentives is that motivations are always personal and vary from individual to individual. Consequently, adapting the incentives and rewards to each particular user can enhance significantly the effectiveness of a motivational strategy (Masthoff et al., 2014).

3.2. Self-Disclosure Biases

As mentioned in section 1, deceivers exploit cognitive and motivational biases that contribute to online self-disclosure to shape their attacks. Hence, determining these biases and how they could be leveraged for malicious personalization is key for maximizing the success and efficiency of an attack. In general, self-disclosure biases have been investigated extensively in psychology through the lens of different theories and behavioral frameworks (Ellison et al., 2007; Steinfield et al., 2008; Stutzman et al., 2011). For instance, studies based on the *use and gratification theory* (McGuire, 1974) have focus on identifying adoption patterns among users of SNSs. That is, they analyse the psychological benefits of engaging in these platforms and sharing information across them (Min and Kim, 2015). In sum, these studies suggest that intrinsic factors like self-promotion (Mehdizadeh, 2010), impression management (Krämer and Winter, 2008), and social capital (Steinfield et al., 2008) may affect users' online behavior. Furthermore, factors like altruism (e.g., provide useful information to help friends) and group joy (e.g., exchange information while interacting in networked games) were also shown to influence people's information-sharing decisions in SNSs (Fu et al., 2017).

Other studies have focused on explaining people's information-sharing behavior through the lens of the *privacy-calculus* (Li et al., 2010; Dienlin and Metzger, 2016; Trepte et al., 2017). That is, they examine how people assess and weigh the costs and benefits of revealing private information when interacting in SNSs. Under this framework, people are expected to open their privacy boundaries (i.e., share more information about themselves) if they outweigh the expected benefits of sharing personal information over their privacy concerns (Laufer and Wolfe, 1977; Culnan and Armstrong, 1999). However, it has been shown that users not always enumerate and evaluate all these costs and benefits in a rational and objective way (Min and Kim, 2015; Trepte et al., 2017). Moreover, it is sometimes hard for regular users to anticipate the consequences of their information-sharing actions, and therefore to make sound privacy decisions (Wang et al., 2011). Hence, factors, such as low levels of literacy and privacy awareness can lead users to disclose information in SNSs which they later regret.

In addition to individual predispositions and cognitive biases, research has also addressed the role of the social context in people's information-sharing behavior (Acquisti and Gross, 2006; Lewis et al., 2008; Cheung et al., 2015; Choi et al., 2018). Overall, this view posits that people often behave in what they believe to be socially accepted ways in order to gain certain benefits so as to avoid social punishment or disapproval. Such socially-compliant decisions are normally made when users lack objective means to evaluate their own behavior (Cialdini and Goldstein, 2004). Social influence has been shown to be a critical factor that determines not only people's engagement in SNSs, but also their privacy behavior within these platforms (Cheung et al., 2015). Particularly, studies have shown that users tend to disclose information about themselves to comply with their peers' expectations (Cheung et al., 2015). Furthermore, they sometimes engage in self-disclosure activities to avoid isolation and, in some cases, to reduce the chances of being stigmatized by others. This last one has been observed in dating apps like *Grindr* in which users include their HIV status as part of their profile to increase their chances of finding a partner (Warner et al., 2018).

4. METHOD

All in all, user's information-sharing behavior is often influenced by their individual motivations and cognitive biases. Likewise, such a behavior can be fostered and guided through personalized incentive mechanisms embedded in the design of information systems. These incentives, when used by deceivers, can be seen as a case of *malicious personalization* in which users are misguided to disclose their private information to others with harmful intentions. In order to understand which cognitive and motivational biases are likely to be exploited by deceivers in SNSs, we have elaborated an online survey about people's willingness to share personal data under different incentives. In this section, the design of such survey is introduced together with the sampling approach.

4.1. Survey Design

To investigate the role of self-disclosure biases in malicious personalization we followed a scenario-based approach. Particularly, participants were asked to indicate their willingness to share pieces of private information under different scenarios. Each scenario represented a situation in which information is asked for apparently harmless purposes (like in deceptive attacks). In total 8 scenarios were included, one for each of the following information categories:

i *Identity*: comprises of identifying information about the users (e.g., name and address).
ii *Social network*: covers information about the social circle and shared content (e.g., friends list and posts).
iii *Health*: includes physical and health related information (e.g., physical condition).
iv *Finances*: encompasses income/expenses and other financial information (e.g., credit card).
v *Education and occupation*: contains information that essentially forms an online résumé (e.g., education level and work experience).
vi *Beliefs*: covers various personal beliefs and points of view (e.g., political and religious views).
vii *Travels*: consists of information about visited locations (e.g., trips to cities and landmarks).
viii *Geolocation*: includes geolocation data (e.g., travels and current GPS position).

For instance, the following scenario was elaborated for the "health" category:

"You start using a fitness tracker/wearable to improve your jogging workout and control your performance. The device app wishes to collect information including your frequent trails, pace, and burnt calories to elaborate a fitness routine for beginners and, thereby, encourage other people to start a healthy lifestyle"

As already mentioned, cognitive and motivational biases may guide user's privacy decisions. On the other hand, deceivers often exploit such biases to manipulate and misguide their victims. Hence, we included for each scenario a set of statements related to the following biases:

- *Financial gain:* The disclosure of personal information is motivated by a cash-equivalent reward, such as money, gifts and discount vouchers (Taylor et al., 2009). This bias could be exploited through a spear-phishing email that says *"We are pleased to announce that employees have the right to get a 50% discount on all of our online products"* and redirects the victim to a phishing page that requests her organizational credentials (i.e., ID and password) to access the discount prize.
- *Personal gain:* The user is motivated to share personal information for a reward that has no cash-equivalent value (Taylor et al., 2009). Such a reward may consist of personalized assistance, customization or any other benefit prized by the user. This bias could be exploited using a spear-phishing email that says *"This is your last chance to get a free premium account at Netflix!"* and asking the organizational credentials of the victim as the required information for the registration.
- *Moral gain (altruism):* The user discloses private information to help others without the expectation of a (not) cash-equivalent reward (Ma and Chan, 2014). For instance, achieving a sense of satisfaction after supporting another user who suffers from the same health condition (Chung, 2014). A deceiver may take advantage of this bias by impersonating a member of a prominent NGO through a fake account and asking to sign a fake petition related to a humanitarian cause.
- *Social compliance:* The users' privacy decisions are influenced by their social context (Cialdini and Goldstein, 2004). Thus, they are more willing to disclose personal information if members of their social circle are already doing it. A deceiver may exploit this bias by asking the victim to answer a fake survey or accessing a non-secure link on behalf of the victim's friends, family or acquaintances.

TABLE 1 | Self-disclosure biases defined for the "health" scenario.

Self-disclosure bias	Survey statement
Financial gain	*"If on exchange I would get a voucher for buying sport clothes, then I would share this data"*
Personal gain	*"If this would grant me access to premium features of the app, then I would allow the app to collect this information"*
Moral gain	*"Since this can help others to develop healthy habits, I would share this information without anything on exchange"*
Social compliance	*"I would share this information with the device if other users start contributing"*
Unawareness	*"I am fine with sharing this information since it is usually collected in an anonymous way"*
Apathy	*"I would give access to this information since these devices are already collecting it for other purposes anyway"*

- *Unawareness:* The user is not able to foresee the (potential) negative consequences of sharing personal information. Hence, the benefits of disclosing such information outweigh the user's underestimated costs (Wang et al., 2011). A deceiver may exploit this bias by claiming to be working in the same company as the victim (e.g., in the IT department) and asking her to start putting confidential information in a non-secure cloud system.
- *Apathy:* The user perceives privacy violations as inevitable and control over personal data as already lost (Hargittai and Marwick, 2016). Such a feeling of resignation drives the user to outweigh the costs of sharing personal information over its potential benefits. A mobile app containing Malware could exploit this bias by simply asking the user to grant full permissions over the phone's GPS location or its photo gallery.

For instance, the *personal gain* statement for the "health" scenario was defined as *"If this would grant me access to premium features of the app, then I would allow the app to collect this information,"* and the corresponding *financial gain* statement as *"If on exchange I would get a voucher for buying sport clothes, then I would share this data."* To evaluate participants' willingness to disclose personal formation, we asked them to indicate to which extent they agree with each of these statements (**Table 1**). For this, a 6-point Likert scale was used were 1 corresponds to "strongly disagree" and 6 to "strongly agree."

Prior to the assessment of the scenarios, participants were asked to answer some questions about their usage of SNSs. Particularly, they were asked (i) how much time do they spend in these platforms, (ii) if they inform themselves about the privacy policies of SNSs, and (iii) if their profile information is made public to others. Participants were also asked to indicate their willingness to sell their private information to SNSs and the value they would assign to different data types. In particular, how cheap/expensive they would sell the information involved in the scenarios they had to evaluate afterwards (i.e., identity, social network, health, finances, education and occupation, beliefs, travels, and geolocation). Specifically, users rated each information category using a 6-point Likert scale where 1 corresponds to "very cheap" and 6 to "very expensive."

4.2. Population and Sampling

The survey was conducted in August of 2019 through Amazon's Mechanical Turk[2] (Mturk), a crowdsourcing marketplace where *requesters* can allocate Human-Intelligence Tasks (HITs) to be completed by the platform's *workers* (Paolacci et al., 2010). Mturk has become a popular platform for researchers to conduct experiments with human subjects particularly in the areas of usable privacy and security (Kelley, 2010). Our HIT was the survey described in section 4.1 and workers were required to have a HIT approval rate ≥95% and a number of approved HITs ≥ 1,000, as it is recommended for this type of task[3]. A remuneration of $1.25 was offered to each worker/participant considering an average completion time of 18 min per survey and the payment standards of the Mturk community. A total of 349 responses from participants of the United States and Canada was considered for the analysis and three were rejected. **Table 2** shows the self-reported demographic characteristics of the study sample.

5. RESULTS AND FINDINGS

Following, we summarize the results of our online survey[4]. Particularly, we analyse how users assess the value of particular pieces of personal information and compare it against their willingness to disclose them under the influence of cognitive and motivational biases (as described in section 4.1). For this, descriptive metrics were elaborated to identify the most reported biases for each scenario. Moreover, a correlation analysis was conducted to investigate relations between survey items. Particularly, to identify correlations between people's willingness to share their personal data and the value they assign to them.

5.1. Cognitive and Motivational Biases

Figure 1 summarizes the participants' assessment of the proposed scenarios. Particularly, their average willingness to share personal data on each specific scenario. As already mentioned, a scenario involves specific type of information and proposes a set of statements related to cognitive and motivational self-disclosure biases. For instance, one can observe that *compliance* and *apathy* are the weakest biases in the scenario concerning financial information. Moreover, together with *moral gain*, *financial gain*, and *unawareness*, have the lowest score across all the scenarios. As **Figure 2** illustrates, the average value assigned to financial data is the highest of all (M = 5.18

[2] Mturk —www.mturk.com

[3] Tips for Academic Requesters on Mturk—http://turkrequesters.blogspot.com/2012/09/tips-for-academic-requesters-on-mturk.html (accessed 07/09/2019).

[4] Survey data is available as **Supplementary Material**.

TABLE 2 | Demographic characteristics of the studied sample.

Demographic	Ranges	Frequency	Responses (%)
Age	18–25 years	13	3.7
	26–35 years	149	42.7
	36–45 years	107	30.7
	46–55 years	46	13.2
	<56 years	34	9.7
Gender	Male	183	52.4
	Female	163	46.7
	Prefer not say	2	0.6
	Non-binary	1	0.3
Occupation	Employed full time	233	66.8
	Employed part time	27	7.7
	Home maker	13	3.7
	Retired	8	2.3
	Self employed	51	14.6
	Student	5	1.4
	Unable to work	4	1.1
	Unemployed	8	2.3
Education	Associate degree	45	12.9
	Bachelor degree	148	42.4
	Doctorate	4	1.1
	High school degree	37	10.6
	Less than high school	2	0.6
	Master degree	42	12
	Professional degree	6	1.7
	Some college, no degree	65	18.6

± 1.139). Hence, this proposes (in principle) that information of high value is less likely to be shared by the users in the context of a deceptive attack. However, reported intentions of sharing other highly-valuable data types like health (M = 5.16 ± 1.211) and identity (M = 5.18 ± 1.139) is high in comparison to other information categories. Furthermore, the statements corresponding to *unawareness* and *apathy* have their highest values on the "health" scenario.

Among all the biases, *personal gain* has its highest peak in the "beliefs" scenario and its second highest in the one of "travels." Moreover, as shown in **Figure 2**, the data corresponding to "beliefs" together with the one of "travels" were reported by the participants as the ones with the lowest value (beliefs: M = 4.07 ± 1.454; travels: M = 3.65 ± 1.51). This suggests, in principle, that *personal gain* can be an influential factor when users are asked for data with a relative low value. However, *personal gain* was also the bias with the highest average score within the "finances" scenario being financial information the one with the highest value. Moreover, this is also the case for the scenarios corresponding to "identity," "social network," "occupation and education," and "travels." Hence, *personal gain* seems to be, in general, the strongest motivation across all the proposed scenarios with the exception of "geolocation" and "health" whose peak correspond to *financial gain* and *apathy*, respectively. On the other hand, *compliance* was the bias with the lowest average score except for the scenarios corresponding to "geolocation" and "social network" in which *moral gain* was rated as the lowest. Likewise, *financial gain* was the bias with the lowest average score in the "education and occupation" scenario.

5.2. Willingness to Share Data

To further investigate users' cognitive and motivational biases when disclosing personal information, we ran an ordinal logistic regression (**Table 3**) which is a widely used method for analysing correlations between Likert items (O'Connell, 2006). For this, the willingness to disclose personal information was defined as the dependent variable and the value of such information as the predictor ("data value"). Therefore, for the eight scenarios/data-types and the six self-disclosure biases, a total of 48 regression analysis were conducted. In addition, the survey items corresponding to (a) having a public profile ("public profile"), and (b) being aware of the privacy policies of SNSs ("policy-aware") were used as control variables.

Table 3 shows the ordered log-odds (B) of the predictors for each bias and disclosure scenario. For instance, one can observe that the log-odds for the reported value of "identity" data is B=-0.390 when the bias is *financial gain*. This means that, for this particular bias, the likelihood of disclosing identity data decreases around $|(e^{-0.39} - 1) * 100| = 32.29\%$ as its value (i.e., the value assigned to "identity" data) increases in one unit. Likewise, this likelihood increases around $|(e^{0.548} - 1) * 100| = 72.98\%$ for those who reported having a public SNS profile. However, there is no statistical significance in relation to the participant's extent of awareness on SNSs' privacy policies.

In general, we observe that, independently of the data type and self-disclosure bias, there is no statistical significance between participants' policy awareness and their reported willingness to disclose personal information. However, having a public SNS profile has shown to have a connection with the reported self-disclosure motivations and cognitive biases. For instance, for biases like *financial* and *moral gain*, the likelihood of disclosing identity, social network and geolocation data increases more than 65% as the survey item "public profile" increases in one unit. This is also the case of *personal gain* and *unawareness* for information related to social network and geolocation, respectively. Furthermore, for *apathy*, the chances of revealing data related to education, identity, and travels rise about 60% per unit of increase in "public profile." Nevertheless, this probability goes bellow 35% in the case of financial information for all the biases. This in principle could be related to the high value assigned to this type of information. However, our sample lacks statistical significance to support this hypothesis. Moreover, "data value" has, in general, very low statistical significance or B-values across the different scenarios and self-disclosure biases.

6. DISCUSSION

Overall, the results of our survey show that self-disclosure biases can vary when people are asked to reveal particular data types. Moreover, a correlation was observed between participants' willingness to reveal personal data and having a public SNSs profile. However, we could not identify correlations for the value

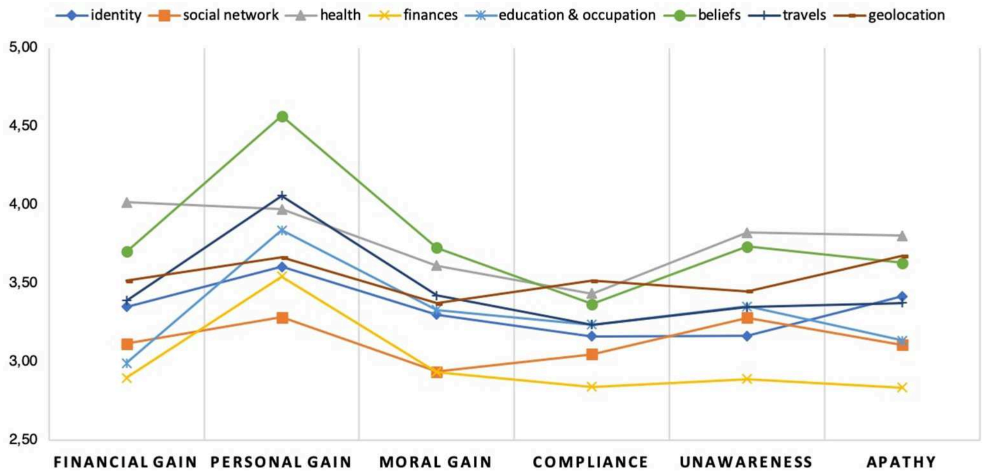


FIGURE 1 | Users' reported cognitive and motivational biases for each scenario (item average).

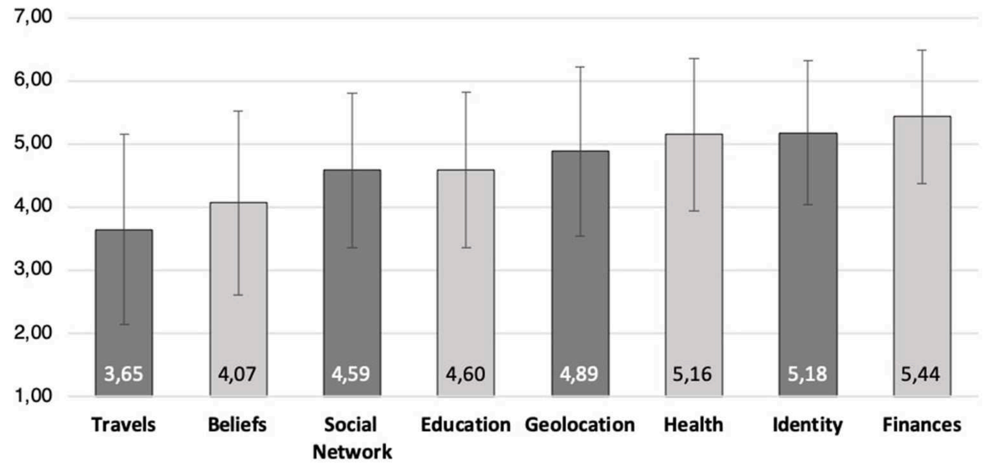


FIGURE 2 | User's reported value for each data type (item average).

participants assign to particular pieces of information, nor their reported awareness level on privacy policies. In the following subsections we discuss the limitations of our approach and elaborate a set of countermeasures based in our findings. The purpose of such countermeasures is to raise awareness among the users of SNSs regarding the potential consequences of revealing private information to deceivers.

6.1. Countermeasures

In order to elaborate deception countermeasures, we first analyse current state-of-the art approaches. Hence, methods and techniques for detecting fake accounts and deceptive messages are discussed in section 6.1.1, and countermeasures are introduced in section 6.1.2. Particularly, the latter section highlights how the findings presented in section 5 can be utilized for the development of personalized risk awareness mechanisms which combine existing approaches together with persuasive technologies.

6.1.1. Current Approaches

Scholars have introduced different strategies to identify deceptive messages and fake accounts in SNSs (Briscoe et al., 2014; Alowibdi et al., 2015; Mulamba et al., 2018; van der Walt et al., 2018). For instance, Briscoe et al. (2014) developed a machine learning model that can detect if a text message sent over a SNS communication channel (e.g., post, tweet, or instant message) is truthful or deceptive. For this, the model uses linguistic cues like the average sentence length, complexity, and sentiment as predictors of deception. On the other hand, Alowibdi et al. (2015) developed a classifier capable to identify inconsistencies in Twitter profiles based on a set of deception indicators (e.g., profile layout colors, first name, and user-name). Particularly, such classifier can detect gender or location inconsistencies in a profile and, thereby, classify its corresponding account as fake. In line with this, van der Walt et al. (2018) followed a similar approach to flag deceptive accounts but using additional predictors, such as tweets geo-tags, name length, and friends/followers ratio.

TABLE 3 | Logistic regression results: ordered log-odds (B) of disclosing personal data on a deceptive scenario.

Bias		Identity	Social network	Health	Geolocation	Travels	Beliefs	Education	Finances
Financial gain	Data value	−0.390***	−0.237*	−0.139†	−0.186*	−0.159*	−0, 150*	−0.093	0.006
	Public profile	0.548***	0.537***	0.334***	0.510***	0.553***	0.409***	0.501***	0.235***
	Policy-aware	0.085	−0.005	−0.054	0.163*	0.235*	0.123	0.071	0.127†
Personal gain	Data value	−0.254*	−0.227*	−0.086	−0.113	−0.224***	−0.288***	−0.207**	0.026
	Public profile	0.317***	0.520***	0.372***	0.339***	0.296***	0.065	0.194***	0.182*
	Policy-aware	0.021	−0.042	−0.090	0.019	0.052	−0.081	−0.021	0.063
Moral gain	Data value	−0.321***	−0.187*	−0.228***	−0.140†	−0.218***	−0.158*	−0.116	−0.082
	Public profile	0.524***	0.546***	0.360***	0.501***	0.0439***	0.410***	0.421***	0.269***
	Policy-aware	0.088	0.116	−0.009	0.144†	0.194*	0.139†	0.063	0.025
Social compliance	Data value	−0.214*	−0.168*	−0.210**	−0.158*	−0.179†	−0.191**	−0.122	0.125
	Public profile	0.546***	0.501***	0.426***	0.486***	0.481***	0.527***	0.459***	0.201***
	Policy-aware	0.098	0.041	0.121	0.112	0.210†	0.201**	0.093	0.098
Unawareness	Data value	−0.332***	−0.203**	−0.081	−0.173*	−0.186**	−0.208**	−0.082	−0.045
	Public profile	0.432***	0.368***	0.431***	0.508***	0.375***	0.301***	0.424***	0.301***
	Policy-aware	−0.071	0.053	0.01	0.196*	0.042	−0.003	−0.011	0.038
Apathy	Data value	−0.275**	−0.262***	−0.186*	−0.091	−0.156*	−0.201**	−0.102	−0.104
	Public profile	0.537***	0.468***	0.404***	0.453***	0.490***	0.439***	0.479***	0.241***
	Policy-aware	0.065	0.033	0.049	0.094	0.119	0.125	0.117	0.065

*†0.05 $< p \leq 0.10$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; $\beta =$ 95%.*

Detecting deceptive accounts and messages is a first attempt on safeguarding the users from harmful online experiences. Furthermore, it is a major step toward ensuring safer interactions through SNSs. However, attacks are getting more sophisticated and, as we can see from the results of our survey, people can be misled to reveal personal information when incentives and motivational biases outweigh their privacy concerns. This demands more effective awareness tools as these instruments play a key role in supporting users when making online privacy decisions. For instance, Díaz Ferreyra et al. (2019) propose the use of *risk patterns* to alert users when they are about to disclose private information inside social media posts. However, to the best of our knowledge, not many efforts have been made on informing the users about the risks of disclosing personal information to deceivers. Particularly, on developing technologies that alert users when they are about to reveal personal data to an attacker.

6.1.2. Personalized Risk Awareness

Overall, current advances on privacy awareness can provide a suitable framework for developing countermeasures against online deception (Petkos et al., 2015; Díaz Ferreyra et al., 2017; De and Le Métayer, 2018). For example, using risk patterns similar to the ones introduced by Díaz Ferreyra et al. (2017) one could define the pre- and post-conditions of a deceptive scenario as a triple *<PI, Deceiver, UIN>* where *PI* corresponds to private information, *Deceiver* to a set of deception queues, and *UIN* to an unwanted incident. Under this representation, the unwanted incident *UIN* corresponds to the *post-condition* of a deceptive attack and revealing the information *PI* to a user with *Deceptive* characteristics to the *pre-condition*. This would allow us, for instance, to represent a scenario in which identity theft (*UIN*) occurs after a user reveals her username and password (*PI*) to another user whose account has been flagged as potentially deceptive (*Deceiver*). Furthermore, a collection of well-known deceptive scenarios expressed in this format could serve the generation of warning messages when the *pre-condition* of one or more patterns is satisfied. For example, showing a pop-up message like *"It seems you are about to reveal <PI> to a user who may be a deceiver. This could derive in a case of <UIN>"* and replacing the place-holders *<PI>* and *<UIN>* with the values defined in the corresponding pattern. This strategy is similar to the one employed by Intelligent Tutoring Systems which are used in learning environments to provide personalized instructional content to students (Díaz Ferreyra, 2019).

The use of interventions (i.e., warning messages or suggestions) is a promising approach for nudging users' privacy behavior (Acquisti et al., 2017). However, it has also been shown that such interventions may result annoying for users with low privacy concerns (Wang et al., 2013). Hence, warnings should be aligned somehow with the privacy goals and expectations of each individual user. In other words, privacy-awareness mechanisms should incorporate adaptivity principles into their design to better engage with their users (Díaz Ferreyra et al., 2019). One of the findings that could contribute in the design of adaptive awareness mechanisms is the one related to the users' profile visibility. Particularly, the *frequency* and *content* of interventions could be tailored using the visibility of the user's profile as an adaptation variable. Moreover, it could be used in combination with the users' privacy attitudes (Ghazinour et al., 2013), risk aversion (Díaz Ferreyra et al., 2019), and digital literacy (Wisniewski et al., 2017) which have already been proposed as variables of adaptation.

On the other hand, the results of our survey also suggest that the influence of self-disclosure biases may vary among users of SNSs. That is, whereas a particular bias can drive a user to disclose her data to a deceiver, the same bias may not influence the behavior of another user under a deceptive attack. Hence, different privacy-awareness strategies may be necessary to deal with the effects of different self-disclosure biases. This could be done, for instance, by framing the *style* of the interventions according to the bias they are addressing (Kaptein et al., 2012). Particularly, interventions may adopt a more *authoritarian* style (e.g., *"Rethink what you are going to provide. Privacy researchers from Harvard University identify such information as highly sensitive!"*) or a more *consensual* one (e.g., *"Everybody agrees: Providing sensitive information can result in privacy risks!"*) depending on the bias they try to counteract (Schäwel and Krämer, 2018). For instance, for users whose more salient bias is *personal gain*, a more authoritarian style could persuade them better than a consensual one. Conversely, for those motivated mainly by *social compliance*, a consensual style may be the most adequate. Besides, warnings could incorporate additional information related to privacy protection mechanisms (e.g., how to block or report a user) to counteract the effect of *apathy*. Furthermore, interventions could also provide links to relevant news and media articles about deception to target *unawareness* or *moral gain* (De and Le Métayer, 2018). In the case of *financial gain*, incorporating information about the value of data together with reputation queues of the data requester may be a good strategy to promote a safer privacy behavior.

6.2. Limitations

Although the approach employed in this work has yielded interesting results, there are some limitations that should be acknowledged. First of all, our results are based on hypothetical self-disclosure scenarios which were evaluated by the participants of our survey. This approach does not ensure that, in a real case scenario, their behavior would be consistent with what they have reported. Likewise, the statements corresponding to the cognitive and motivational biases we defined should be elaborated further, especially in the form of validated Likert scales. On the other hand, using Mturk for conducting online surveys supposes a loss of control over the experimental setting on a large extent (Kittur et al., 2008; Paolacci et al., 2010). In particular, participants may get distracted in their physical environment and, thereby, compromise the quality of their answers. Furthermore, workers sometimes provide fast or nonsense answers in order to make more money in less time. Nevertheless, it has been shown that the Mturk platform can provide results as relevant as those from traditional survey methods (Paolacci et al., 2010). This can be achieved by applying a number of good practices, such as controlling the time workers actually spend in the task or filter out workers with a low HIT approval rate (Amazon, 2011; Oh and Wang, 2012). Such practices were followed to ensure good quality results.

7. CONCLUSIONS AND FUTURE WORK

Safeguarding people's private information is extremely important for the welfare of modern societies. However, increasing the security levels around such information is not enough since nowadays it is possible to monitor and analyse people through their SNS profiles. This makes cyber-attacks very easy to personalize according to what hackers may find about their victims in these online platforms. It is not a secret that, for example, identity theft affects millions of people a year costing victims countless hours and money in identity recovery and repair. The much-publicized Equifax scandal that broke out in September 2017—after the personal information of as many as 143 million Americans had been compromised (and an untold number of Canadians and Brits)—has resulted in the recent resignation of the Equifax CEO. Even Hollywood makes films about cases of extreme lack of privacy, such as *The Circle*, and about personalization of phishing attacks, such as *CSI: Cyber*.

In sum, we need to provide a better future for the next generation of Internet users since it will be born in an age in which privacy may appear as an anomaly. However, people will remain susceptible to manipulation and privacy risks unless coordinated actions between developers of media technologies, users, government, and the civil society are jointly taken. This work has explored the exploitable biases for malicious personalization in SNSs and elaborated countermeasures which incorporate current advances in risk awareness, personalization and persuasive technologies. We believe that such countermeasures are a promising approach for engaging users of SNSs (specially teenagers) in a sustained privacy-learning process. Moreover, the premise of such countermeasures is not banning people from sharing status updates, photos and networking, but to support them in their individual privacy decisions. This would not only increase their levels of risk awareness but also allow them to disclose private information at their own responsibility.

As mentioned throughout this work, deceptive attacks are hard to identify since deceivers employ different strategies (i.e., motivations and incentives) to influence and mislead their victims. Moreover, such attacks can be crafted and personalized to the particular self-disclosure biases of the targeted victim in order to maximize their damage. Hence, understanding the cognitive and motivational biases exploited by deceivers is necessary for shaping privacy-preserving technologies to protect the users. The results of this work suggest that, in principle, the effect of each bias vary from individual to individual. Therefore, technical countermeasures as well as training and awareness programs should be personalized according to the biases that are more exploitable for each particular user. Moreover, the use of risk communication strategies is a promising approach for designing personalized countermeasures and will be investigated in further publications.

One of the most salient findings of this work is the relation between users' profile visibility and their willingness to share private information under a deceptive attack. Specifically, it was observed that participants who reported having a public profile were more willing to disclose personal data in a deceptive scenario. Therefore, profile visibility is proposed as a potentially significant adaptation variable for deception countermeasures. However, recent research in online self-disclosure has found no differences in the self-disclosure practices of users with a public SNS profile and those with a private one (Gruzd and Hernández-García, 2018). Nevertheless, that study did not take into consideration the influence that incentive mechanisms together with cognitive and motivational biases may have on users' privacy practices. Hence, we intend to research this point in more detail, in order to further corroborate our results.

Another aspect that should be analyzed in more detail are the cultural factors that may influence people's privacy decisions. Particularly, the results of this work are based on a sample consisting of Americans and Canadians which, according to the Hoftede's taxonomy, are *individualistic* societies (Li et al., 2017). That is, they tend to care more of themselves and their inner circle, and exhibit a behavior which is mainly driven by individual achievements. Conversely, in *collectivist* societies, such as Mexico or Spain, people often reflect on the consequences that their actions may have on others; particularly on the members of their social context (e.g., extended families, clans, or organizations) (Hofstede, 2011). Thus, some of the results presented in this work may be closely connected to the cultural background of the survey participants. For instance, the prevalence of "personal gain" in most of the scenarios may be due to the individualistic nature of the sample among other cultural factors. Hence, future research will investigate further the effects of the social context on the motivations and cognitive biases which are frequently exploited by deceptive attackers.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of the University of Montreal. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The study was outlined and conceived by EA and HH. ND organized the dataset, performed the statistical analysis and wrote the first draft of the manuscript. All authors contributed to the design of the study, manuscript revision, read, and approved the submitted version.

FUNDING

This work was partially supported by Canada's Natural Sciences and Engineering Research Council (NSERC) and the H2020 European Project No. 787034 PDP4E: Privacy and Data Protection Methods for Engineering.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2019.00026/full#supplementary-material

REFERENCES

Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., et al. (2017). Nudges for privacy and security: understanding and assisting users' choices online. *ACM Comput. Surv.* 50:44. doi: 10.1145/3054926

Acquisti, A., and Gross, R. (2006). "Imagined communities: awareness, information sharing, and privacy on the Facebook," in *Privacy Enhancing Technologies*, eds G. Danezis and P. Golle (Berlin; Heidelberg: Springer), 36–58.

Aïmeur, E., Hage, H., and Amri, S. (2018). "The scourge of online deception in social networks," in *Proceedings of the 2018 Annual Conference on Computational Science and Computational Intelligence (CSCI' 18)*, eds H. R. Arabnia, L. Deligiannidis, F. G. Tinetti, and Q.-N. Tran (Las Vegas, NV: IEEE Computer Society), 1266–1271.

Aïmeur, E., and Sahnoune, Z. (2019). Privacy, trust, and manipulation in online relationships. *J. Technol. Hum. Serv.* doi: 10.1080/15228835.2019.1610140. [Epub ahead of print].

Albladi, S., and Weir, G. R. (2016). "Vulnerability to social engineering in social networks: a proposed user-centric framework," in *2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF)* (Vancouver, BC: IEEE), 1–6.

Alowibdi, J. S., Buy, U. A., Yu, P. S., Ghani, S., and Mokbel, M. (2015). Deception detection in Twitter. *Social Netw. Anal. Mining* 5:32. doi: 10.1007/s13278-015-0273-1

Amazon (2011). *Mechanical Turk: Requester Best Practices Guide*. Technical report, Amazon Inc.

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York, NY: Harper Collins.

Berkovsky, S., Freyne, J., and Oinas-Kukkonen, H. (2012). Influencing individually: fusing personalization and persuasion. *ACM Trans. Interact. Intell. Syst.* 2:9. doi: 10.1145/2209310.2209312

Boyd, D. (2010). "Social network sites as networked publics: affordances, dynamics, and implications," in *A Networked Self,* ed Z. Papacharissi (New York, NY: Routledge), 47–66.

Briscoe, E. J., Appling, D. S., and Hayes, H. (2014). "Cues to deception in social media communications," in *47th Annual Hawaii International Conference on System Sciences (HICSS)* (Hawaii: IEEE), 1435–1443.

Burke, M., Marlow, C., and Lento, T. (2009). "Feed me: motivating newcomer contribution in social network sites," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09* (Boston, MA: ACM), 945–954.

Chang, D., Krupka, E. L., Adar, E., and Acquisti, A. (2016). "Engineering information disclosure: norm shaping designs," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '06* (San Jose, CA: ACM), 587–597.

Cheung, C., Lee, Z. W. Y., and Chan, T. K. H. (2015). Self-disclosure in social networking sites: the role of perceived cost, perceived benefits and social influence. *Internet Res.* 25, 279–299. doi: 10.1108/IntR-09-2013-0192

Choi, H. S., Kwak, C., Lee, J., and Lee, H. (2018). "Too crowded to disclose! Exploring the relationship between online crowdedness and self-disclosure," in *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)* (Hawaii), 3708–371.

Chung, J. E. (2014). Social networking in online support groups for health: how online social networking benefits patients. *J. Health Commun.* 19, 639–659. doi: 10.1080/10810730.2012.757396

Cialdini, R. B., and Goldstein, N. J. (2004). Social influence: compliance and conformity. *Annu. Rev. Psychol.* 55, 591–621. doi: 10.1146/annurev.psych.55.090902.142015

Conti, G., and Sobiesk, E. (2009). Malicious interfaces and personalization's uninviting future. *IEEE Secur. Privacy* 7, 64–67. doi: 10.1109/MSP.2009.71

Culnan, M. J., and Armstrong, P. K. (1999). Information privacy concerns, procedural fairness, and impersonal trust: an empirical investigation. *Organ. Sci.* 10, 104–115.

De, S. J., and Le Métayer, D. (2018). "Privacy risk analysis to enable informed privacy settings," in *2018 IEEE European Symposium on Security and Privacy Workshops* (London, UK), 95–102.

Díaz Ferreyra, N. E. (2019). *Instructional awareness: a user-centred approach for risk communication in social network sites* (Ph.D. thesis), DuEPublico: Duisburg-Essen Publications, University of Duisburg-Essen, Duisburg, Germany.

Díaz Ferreyra, N. E., Meis, R., and Heisel, M. (2017). "Online self-disclosure: from users' regrets to instructional awareness," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, eds A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl (Reggio Calabria: Springer International Publishing), 83–102.

Díaz Ferreyra, N. E., Meis, R., and Heisel, M. (2019). "Learning from online regrets: from deleted posts to risk awareness in social network sites," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (Larnaca: ACM), 117–125.

Dienlin, T., and Metzger, M. J. (2016). An extended privacy calculus model for SNSs: analyzing self-disclosure and self-withdrawal in a representative U.S. sample. *J. Comput. Mediat. Commun.* 21, 368–383. doi: 10.1111/jcc4.12163

Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook "friends": social capital and college students' use of online social network sites. *J. Comput. Mediat. Commun.* 12, 1143–1168. doi: 10.1111/j.1083-6101.2007.00367.x

Fu, P.-W., Wu, C.-C., and Cho, Y.-J. (2017). What makes users share content on Facebook? Compatibility among psychological incentive, social capital focus, and content type. *Comput. Hum. Behav.* 67, 23–32. doi: 10.1016/j.chb.2016.10.010

Ghazinour, K., Matwin, S., and Sokolova, M. (2013). "Monitoring and recommending privacy settings in social networks," in *Proceedings of the Joint EDBT/ICDT 2013 Workshops, EDBT '13* (Genoa: ACM), 164–168.

Gruzd, A., and Hernández-García, Á. (2018). Privacy concerns and self-disclosure in private and public uses of social media. *Cyberpsychol. Behav. Soc. Netw.* 21, 418–428. doi: 10.1089/cyber.2017.0709

Hamari, J., and Koivisto, J. (2013). "Social motivations to use gamification: an empirical study of gamifying exercise," in *Proceedings of the 21st European Conference on Information Systems (ECIS)* (Utrecht).

Hargittai, E., and Marwick, A. (2016). "What can I really do?" Explaining the privacy paradox with online apathy. *Int. J. Commun.* 10:21. doi: 10.5167/uzh-148157

Hofstede, G. (2011). Dimensionalizing cultures: the hofstede model in context. *Online Read. Psychol. Cult.* 2:8. doi: 10.9707/2307-0919.1014

Joinson, A. N. (2008). "Looking at, looking up or keeping up with people? Motives and use of Facebook," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (Florence: ACM), 1027–1036.

Kaptein, M., De Ruyter, B., Markopoulos, P., and Aarts, E. (2012). Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Trans. Interact. Intell. Syst.* 2:10. doi: 10.1145/2209310.2209313

Kelley, P. G. (2010). "Conducting usable privacy & security studies with Amazon's mechanical turk," in *Symposium on Usable Privacy and Security (SOUPS)* (Redmond, WA).

Kittur, A., Chi, E. H., and Suh, B. (2008). "Crowdsourcing user studies with mechanical turk," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence: ACM), 453–456.

Krämer, N. C., and Winter, S. (2008). Impression management 2.0: the relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *J. Media Psychol.* 20, 106–116. doi: 10.1027/1864-1105.20.3.106

Kraut, R. E., and Resnick, P. (2012). *Building Successful Online Communities: Evidence-Based Social Design*. MIT Press.

Krombholz, K., Hobel, H., Huber, M., and Weippl, E. (2015). Advanced social engineering attacks. *J. Inform. Secur. Appl.* 22, 113–122. doi: 10.1016/j.jisa.2014.09.005

Laufer, R. S., and Wolfe, M. (1977). Privacy as a concept and a social issue: a multidimensional developmental theory. *J. Soc. Issues* 33, 22–42.

Lewis, K., Kaufman, J., and Christakis, N. (2008). The taste for privacy: an analysis of college student privacy settings in an online social network. *J. Comput. Mediat. Commun.* 14, 79–100. doi: 10.1111/j.1083-6101.2008.01432.x

Li, H., Sarathy, R., and Xu, H. (2010). Understanding situational online information disclosure as a privacy calculus. *J. Comput. Inform. Syst.* 51, 62–71. doi: 10.1080/08874417.2010.11645450

Li, Y., Kobsa, A., Knijnenburg, B. P., and Nguyen, M.-H. C. (2017). Cross-cultural privacy prediction. *Proc. Privacy Enhancing Technol.* 2017, 113–132. doi: 10.1515/popets-2017-0019

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., et al. (2005). Using social psychology to motivate contributions to online communities. *J. Comput. Mediat. Commun.* 10:JCMC10411. doi: 10.1111/j.1083-6101.2005.tb00273.x

Ma, W. W. K., and Chan, A. (2014). Knowledge sharing and social media: altruism, perceived online attachment motivation, and perceived online relationship commitment. *Comput. Hum. Behav.* 39, 51–58. doi: 10.1016/j.chb.2014.06.015

Masthoff, J., Grasso, F., and Ham, J. (2014). Preface to the special issue on personalization and behavior change. *User Model. User Adapt. Interact.* 24, 345–350. doi: 10.1007/s11257-014-9151-1

McGuire, W. J. (1974). Psychological motives and communication gratification. *Uses Mass Commun. Curr. Perspect. Gratif. Res.* 3, 167–196.

Mehdizadeh, S. (2010). Self-presentation 2.0: narcissism and self-esteem on Facebook. *Cyberpsychol. Behav. Soc. Netw.* 13, 357–364. doi: 10.1089/cyber.2009.0257

Min, J., and Kim, B. (2015). How are people enticed to disclose personal information despite privacy concerns in social network sites? The calculus between benefit and cost. *J. Assoc. Inform. Sci. Technol.* 66, 839–857. doi: 10.1002/asi.23206

Mulamba, D., Ray, I., and Ray, I. (2018). "On Sybil classification in online social networks using only structural features," in *2018 16th Annual Conference on Privacy, Security and Trust (PST)* (Belfast: IEEE), 1–10.

O'Connell, A. A. (2006). *Logistic Regression Models for Ordinal Response Variables*, Vol. 146. Sage.

Oh, J., and Wang, G. (2012). "Evaluating crowdsourcing through Amazon mechanical turk as a technique for conducting music perception experiments," in *Proceedings of the 12th International Conference on Music Perception and Cognition* (Thessaloniki), 1–6.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgm. Decis. Mak.* 5, 411–419. Available online at: https://ssrn.com/abstract=1626226

Penni, J. (2017). The future of online social networks (OSN): a measurement analysis using social media tools and application. *Telem. Inform.* 34, 498–517. doi: 10.1016/j.tele.2016.10.009

Petkos, G., Papadopoulos, S., and Kompatsiaris, Y. (2015). "PScore: a framework for enhancing privacy awareness in online social networks," in *Proceedings of the 10th International Conference on Availability, Reliability and Security, ARES 2015* (Toulouse: IEEE), 592–600.

Schäwel, J., and Krämer, N. (2018). *Do You Really Want to Disclose? Examining Psychological Variables that Influence the Effects of Persuasive Prompts for Reducing Online Privacy Risks*. Frankfurt: Forschungsreferat beim 51. Kongress der Deutschen Gesellschaft für Psychologie (DGPs).

Skinner, B. F. (1969). *Contingencies of Reinforcement*. New York, NY: Appleton-Century-Crofts.

Smith, A., and Anderson, M. (2018). *Social Media Use in 2018*. Pew Research Center: Internet & Technology.

Steinfield, C., Ellison, N. B., and Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. *J. Appl. Dev. Psychol.* 29, 434–445. doi: 10.1016/j.appdev.2008.07.002

Stutzman, F., Capra, R., and Thompson, J. (2011). Factors mediating disclosure in social network sites. *Comput. Hum. Behav.* 27, 590–598. doi: 10.1016/j.chb.2010.10.017

Such, J. M., and Criado, N. (2018). Multiparty privacy in social media. *Commun. ACM* 61, 74–81. doi: 10.1145/3208039

Taylor, D. G., Davis, D. F., and Jillapalli, R. (2009). Privacy concern and online personalization: the moderating effects of information control and compensation. *Electron. Commerce Res.* 9, 203–223. doi: 10.1007/s10660-009-9036-2

Trepte, S., Reinecke, L., Ellison, N. B., Quiring, O., Yao, M. Z., and Ziegele, M. (2017). A cross-cultural perspective on the privacy calculus. *Social Media Soc.* 3, 1–13. doi: 10.1177/2056305116688035

Tsikerdekis, M., and Zeadally, S. (2014). Online deception in social media. *Commun. ACM* 57:72. doi: 10.1145/2629612

van der Walt, E., and Eloff, J. H. (2017). "Identity deception detection on social media platforms," in *Proceedings of the 3rd International Conference on Information Systems Security and Privacy (ICISSP)* (Porto), 573–578.

van der Walt, E., Eloff, J. H. P., and Grobler, J. (2018). Cyber-security: identity deception detection on social media platforms. *Comput. Secur.* 78, 76–89. doi: 10.1016/j.cose.2018.05.015

Vassileva, J. (2012). Motivating participation in social computing applications: a user modeling perspective. *User Model. User Adapt. Interact.* 22, 177–201. doi: 10.1007/s11257-011-9109-5

Vroom, V. H. (1964). *Work and Motivation*, Vol. 54. New York, NY: Wiley.

Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., and Cranor, L. F. (2013). "Privacy nudges for social media: an exploratory Facebook study," in *Proceedings of the 22nd International Conference on World Wide Web* (Rio de Janeiro: ACM), 763–770.

Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., and Cranor, L. F. (2011). ""I regretted the minute I pressed share": a qualitative study of regrets on Facebook," in *Proceedings of the 7th Symposium on Usable Privacy and Security, SOUPS 2011* (Pittsburgh, PA: ACM), 1–16.

Warner, M., Gutmann, A., Sasse, M. A., and Blandford, A. (2018). Privacy unraveling around explicit HIV status disclosure fields in the online geosocial hookup app grindr. *Proc. ACM Hum. Comput. Interact.* 2:181. doi: 10.1145/3274450

Wisniewski, P. J., Knijnenburg, B. P., and Lipford, H. R. (2017). Making privacy personal: profiling social network users to inform privacy education and nudging. *Int. J. Hum. Comput. Stud.* 98, 95–108. doi: 10.1016/j.ijhcs.2016.09.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 17 December 2019
doi: 10.3389/frai.2019.00030

Apps for Mental Health: An Evaluation of Behavior Change Strategies and Recommendations for Future Development

Felwah Alqahtani[1,2]*, *Ghazayil Al Khalifah*[1], *Oladapo Oyebode*[1] *and Rita Orji*[1]

[1] *Computer Science, Dalhousie University, Halifax, NS, Canada,* [2] *Computer Science, King Khalid University, Abha, Saudi Arabia*

Mobile applications have shown promise in supporting people with mental health issues to adopt healthy lifestyles using various persuasive strategies. However, the extent to which mental health apps successfully employ various persuasive strategies remains unknown. Hence, it is important to understand the persuasive strategies integrated into mental health applications (apps) and how they are implemented to promote mental health. This paper aims to achieve three main objectives. First, we review 103 mental health apps and identify distinct persuasive strategies incorporated in them using the Persuasive Systems Design (PSD) model and Behavior Change Techniques (BCTs). We further classify the persuasive strategies based on the type of mental health issues the app is focused on. Second, we reveal the various ways that the persuasive strategies are implemented/operationalized in mental health apps to achieve their intended objectives. Third, we examine the relationship between apps effectiveness (measured by user ratings) and the persuasive strategies employed. To achieve this, two researchers independently downloaded and used all identified apps to identify the persuasive strategies using the PSD model and BCTs. Next, they also examine the various ways that these strategies are implemented in mental health apps. The results show that the apps employed 26 distinct persuasive strategies and a range of 1–10 strategies per app. *Self-monitoring* ($n = 59$), *personalization* ($n = 55$), and *reminder* ($n = 49$) were the most frequently employed strategies. We also found that anxiety, stress, depression, and general mental health issues were the common mental health issues targeted by the apps. Finally, we offer some design recommendations for designing mental health apps based on our findings.

Keywords: persuasive strategies, mental health, mobile application, evaluation, implementation

OPEN ACCESS

Edited by:
Panagiotis Germanakos, SAP SE, Germany

Reviewed by:
Mazida Ahmad, Northern University of Malaysia, Malaysia
Stephan Weibelzahl, Private University of Applied Sciences, Germany

****Correspondence:***
Felwah Alqahtani
felwah.alqahtani@dal.ca

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *11 September 2019*
Accepted: *29 November 2019*
Published: *17 December 2019*

Citation:
Alqahtani F, Al Khalifah G, Oyebode O and Orji R (2019) Apps for Mental Health: An Evaluation of Behavior Change Strategies and Recommendations for Future Development. Front. Artif. Intell. 2:30. doi: 10.3389/frai.2019.00030

INTRODUCTION

Nowadays, mental health issues have become a major public health challenge. People with mental health issues find it difficult achieving their daily tasks such as work and study (Keyes, 2005). As result, many of them are using digital applications to support their mental health and enhance life quality. More than 10,000 mental health and wellness apps are available for download and use (Torous and Roberts, 2017) online. The ubiquitous nature of smartphones and other handheld mobile devices are shaping-up users' lifestyles by adding new aspects to the concept of socializing, accomplishing actions, and creating new habits (Oulasvirta et al., 2012). Therefore, smartphones are attractive platforms for researchers to deliver interventions. Mobile applications (apps) are being used to deliver interventions targeting various health issues (Iacoviello et al., 2017). For mental health issues specifically, Roepke et al. (2015) and Arean et al. (2016) highlighted in their studies that mobile-based mental health intervention made a strong impact on reducing depressed mood. However, they also reported a high rate of drop-out.

By applying various persuasive strategies to reinforce, change, or shape users' behavior and/or attitudes, mental health apps can effectively function as support tools that also motivate and stimulate users to keep on using the apps to achieve better mental health. However, the extent to which available mental health apps successfully employed persuasive strategies and how they implement them in their app to achieve their intended objective remains unknown.

Therefore, this paper aims to achieve three main objectives. First, we review 103 mental health applications and identify distinct persuasive strategies incorporated in them using the Persuasive Systems Design (PSD) model and Behavior Change Techniques (BCTs). We further classify the persuasive strategies based on the type of mental health issues the app is focused on. Second, we reveal the various ways that the persuasive strategies are implemented/operationalized in mental health applications to achieve their intended objectives. Third, we examine whether there is relationship between apps effectiveness (measured by user ratings) and the persuasive strategies employed. To achieve this, two researchers independently downloaded and used all identified apps to identify the persuasive strategies using the PSD model and BCTs. Next, they also examine the various ways that these strategies were implemented in the mental health apps. The results show that the apps employed 26 distinct persuasive strategies and a range of 1–10 strategies per app. Self-monitoring ($n = 59$), personalization ($n = 55$), and reminder ($n = 49$) were the most frequently employed. We also found that anxiety, stress, depression, and general mental health issues were the common mental health issues targeted by the apps. Finally, we offer some design recommendations for designing mental health apps based on our results.

Identifying the persuasive strategies in mental health apps, classifying them based on the type of mental health issues the apps target, and uncovering the relationship between app effectiveness and persuasive strategies employed would be valuable for both researchers and developers working in the mental health domain to inform the design of mental health apps.

BACKGROUND

Interactive systems that are designed to change users' behavior or attitude in an intended way are called Persuasive Systems (PSs) (Fogg, 2009). Persuasive systems are widely used in the health and wellness domains to encourage and help users to change their behaviors and/or attitudes.

According to Fogg's Behavior Model (FBM) (Fogg, 2009), there are three factors that help users to perform their target behavior and/or attitudes. These factors are motivation, ability, and triggers. Increasing these three factors is the main focus of persuasive systems. The aim of FBM is to assist researchers and designers to think more about the target behavior that needs to be changed and understand how to design persuasive systems to achieve the desired outcome (Fogg, 2009).

Over the years, many frameworks and taxonomies exist to help designers of persuasive systems in understanding and deconstructing techniques employed in persuasive systems design. Harjumaa and Oinas-Kukkonen (2009) proposed 28 principles of persuasive system design (PSD) model based on three stages of PS development: (1) understanding the main issue behind PS, (2) analyzing the context of PS, and (3) describing different methods to design system features. These principles were classified into four main categories: primary task support, dialogue support, credibility support, and social support categories. Similarly, Abraham and Michie (2008) developed the Behavior Change Techniques (BCTs) which consists of 26 BCTs taxonomies. This Taxonomy was extended later by Michie et al. (2013) to include 93 BCTs, called Behavior Change Technique Taxonomy.

Many researchers used either the PSD model or BCTs or a combination of both to study and deconstruct the persuasive strategies employed in persuasive systems in many areas including web-based health interventions (Kelders et al., 2012) and mobile health interventions (Almutari and Orji, 2019).

Several studies include conducting a systematic review to identify persuasive strategies implemented in health applications using PSD/BCT. For example, Kelders et al. (2012) conducted a systematic review of web-based health interventions and used PSD model to identify which persuasive strategies were most commonly employed and how they affected adherence to the interventions. Their results show that most web-based persuasive systems employed strategies from the primary task support categories including tunneling, tailoring, reduction, and self-monitoring compared to the strategies in the social support category. However, while social support strategies were less commonly employed in web-based interventions, they show a significant contribution to better adherence. In contrast, primary task support strategies that were mostly implemented in web-based interventions did not show any predictive value for adherence. Kelders et al. stated that using persuasive strategies can demonstrate a significant amount of difference in adherence.

Lehto and Oinas-Kukkonen (2011) focused on identifying the persuasive strategies in web-based alcohol and smoking interventions using PSD model. They found that primary task support strategies such as reduction and self-monitoring were widely employed whereas there was a lack of tailoring, which might mean that the interventions are not targeting a particular audience. Similarly, Crane et al. (2015) reviewed popular alcohol-related apps to identify BCTs and discovered that facilitating self-recording information on the consequences of excessive alcohol use, alongside performance feedback, were these apps' most employed strategies.

Matthews et al. (2016) conducted a systematic review of 20 research papers to describe the use of persuasive strategies on mobile apps promoting physical activity using the PSD model. They found that self-monitoring was the most commonly employed strategy whereas system credibility support category was absent in most reviewed mobile applications. However, credibility support strategies, including surface credibility and expertise, were the highest implemented strategies for chronic arthritis apps, followed by general information and self-monitoring (Geuens et al., 2016).

Almutari and Orji (2019) only examined the 32 papers that implemented social support strategies to understand their effectiveness in encouraging physical activity using PSD. They discovered that competition, social comparison, and cooperation are effective strategies to motivate physical activity. For medication management apps for consumers, a reminder was the highest strategy implemented in those apps, followed by tailoring and self-monitoring (Win et al., 2017).

Additionally, Gardner et al. (2016) focused on identifying strategies employed in sedentary behavior reduction interventions using BCT. The study found that the most frequently observed strategies were setting behavioral goals, providing unspecified forms of social support, instruction on how to perform the behavior and self-monitoring. However, self-monitoring, problem solving, and restructuring the social or physical environment were particularly promising behavior-change strategies.

Chang et al. (2012) conducted a systematic review of persuasive strategies in 12 mental health apps to identify the persuasive strategies that are employed in them using the PSD model. They found that primary task support strategies were the most commonly employed whereas social support strategies were least commonly employed. Overall, they concluded that persuasive strategies were not widely employed in mental health applications. Moreover, Wildeboer et al. (2016) examined the relationship between persuasive strategies, adherence, and the effectiveness of web-based intervention for mental health. Results indicated there is a relationship between the number of persuasive strategies and the intervention's effectiveness. We extend existing work by focusing on persuasive strategies in mental health apps and employing both the BCTs and the PSD framework in our review. First, we identify distinct persuasive strategies incorporated in mental health apps and classify the persuasive strategies based on the type of mental health issues the app is focused on. Second, we reveal the various ways that the persuasive strategies are implemented/operationalized in mental health applications to achieve their intended objectives. Third, we examine the relationship between app effectiveness (measured by user ratings) and the persuasive strategies employed. Finally, we offer some design recommendations that help app developers and health professionals to build more effective support tools for people who experience mental health issues.

METHODS

In this section, we describe the methods used to achieve the study objectives. Specifically, we detailed the app selection criteria and the coding.

Selection of Sample Apps

We searched on the App Store and Google Play using the keywords "mental health," "anxiety," "depression," "mood," "emotions," and "stress." We also searched using various combinations of the keywords joined using the conjunctions "OR" and "AND." The search result revealed the initial list of 437 apps (258 apps from App Store and 179 apps from Google Play). For our analysis, we included apps whose main goal according to the app's description and the demo of the app show that they are targeted at mental health, and apps that have more than five reviews (comments) in total. In other words, apps that fall into any of these categories are excluded: (1) not focused on mental health, (2) had less than five reviews (or comments), or (3) was not in English. In addition, for apps that appeared in both App Store and Google Play, we counted it as one instead of two. After applying the selection criteria, a total of 103 apps remained and eligible for coding (see **Figure 1**). The following information was also extracted for each eligible app: *name, platform* (i.e., iPhone, Android, or both), *developer, date of the last update,* and *price* (i.e., free, fee-based, and free with in-app purchases—where developers provide a free version and a paid version if users want to upgrade or unlock additional features in the app).

Coding Apps for Persuasive Strategies

The aim of the coding process in our study is to assess the number and type of persuasive strategies present in mental health apps. Collected apps were coded using both the Persuasive System Design model (PSD) (Harjumaa and Oinas-Kukkonen, 2009) by Oinas-Kukkonen and Harjuma, and Behavior Change Techniques BCTs (Michie et al., 2013) by Michie and Abraham. We combined and used both the Oinas-Kukkonen and Michie's frameworks to have a comprehensive list of strategies for deconstructing the apps.

To identify the persuasive strategies employed in the apps and their implementations, a subsample apps (and=5) were downloaded and used for 10 days by two researchers to ensure

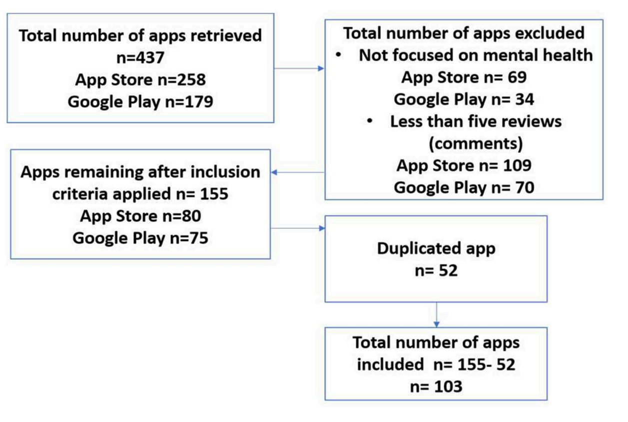


FIGURE 1 | Process of selecting mental health apps.

there is no new strategy revealed during using. After that two researchers independently downloaded and reviewed the 103 apps to identify the persuasive strategies using the PSD model and BCTs. The researchers then met to agree on the initial codes. For any disagreement that arose between the two researchers, a third researcher was involved to mediate and ensure an agreement is reached. The researchers also classified the persuasive strategies based on the type of mental health issues the app is targeting. Moreover, we also identified the various ways that the persuasive strategies are implemented/operationalized in mental health apps to achieve their intended objectives.

Analysis

To analyze our data, we employed some well-known analytical approaches:

- First, we measured the percentage of agreement between two researchers (i.e., before the third researcher was involved). We also calculated interrater reliability, kappa and prevalence and bias adjusted kappa (PABAK).
- Second, we conducted descriptive statistics to obtain the mean of persuasive strategies employed in the apps.
- Third, we employed independent-samples t-tests to compare the mean of persuasive strategies between free and paid apps and between iPhone and Android apps. Apps that have two versions used on both platforms were not included in the t-test analysis.
- Finally, to examine the relationship between the number of persuasive strategies and the effectiveness of apps (as determined by the app ratings), we performed a Pearson's correlation analysis between the number of persuasive strategies and the app's rating.

TABLE 1 | A summary description of 103 mental health apps.

Price	Free (41), fee-based (11%), Free with in app purchases (49%)
Developer	Unknown (16%), Commercial (profit Organization) (69%), Government (7%), NGO (4%), University (5%)
Rating	No rating (7%), 2–2.9 (4%), 3–3.9 (16%), 4–4.9 (69%), 5(4%).
Platform	iPhone (27%), Android (22%), both (50%).

RESULTS

We present the detailed results of our analysis in subsequent subsections. We describe the coding agreement, the persuasive strategies employed, their implementations, the target mental health domain, the relationship between the number of strategies employed and app effectiveness.

Description of Selected Apps

We provide a summary of the app's description in **Table 1**. Approximately half (47%) of the apps had been updated within the past year (2018). More details of the apps can be found in the **Appendix**.

Persuasive Strategies Employed in Mental Health Apps

The results of our analysis show that the percentage of agreement between the two researchers was 86.5%. There was "substantial" agreement: prevalence and bias adjusted kappa (PABAK) = 0.71 (Landis and Koch, 1977; Byrt et al., 1993). Discrepancies were discussed and the coding was refined. Overall, we found 26

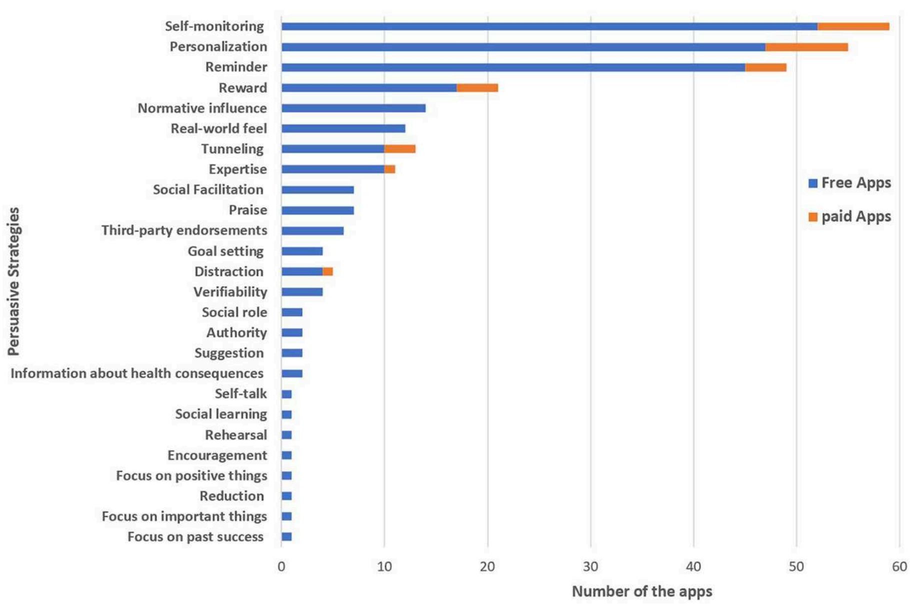


FIGURE 2 | Persuasive strategies employed by mental health apps categorized into free and paid apps.

distinct persuasive strategies present in the mental health apps reviewed. The number of strategies employed in each app varied and ranges between 1 and 10. However, 14 mental health apps did not employ any persuasive strategies. Interestingly, self-monitoring ($n = 59$), personalization ($n = 55$), and reminder ($n = 49$) emerged as the most commonly employed strategies (see **Figure 2**). Moreover, we found other strategies that do not exist in PSD/BCT that were employed in the reviewed mental health apps: *Encouragement, focus on positive things* and *focus on important things*.

The results of our *t*-test show that there was a significant difference in the number of persuasive strategies employed within iPhone apps ($M = 2.61$, $SD = 1.524$) and Android apps ($M= 1.74$, $SD = 1.137$); $t(49) = 2.26$, $p = 0.028$. These results suggest that the number of persuasive strategies employed within iPhone apps is more than Android apps. However, there was no significant difference in the number of persuasive strategies present in free apps ($M = 2.38$, $SD = 2.073$) and paid apps ($M = 2.27$, $SD = 1.104$) apps; $t(101) = 0.868$, $p = 0.388$. These results suggest that the number of persuasive strategies present in free apps is the same in paid apps.

Persuasive Strategies in Other Health Domains

Comparison of our findings to the findings of earlier reviewed health care apps studies reveals that self-monitoring is one of the common strategies that emerge in chronic arthritis apps (Geuens et al., 2016), alcohol reduction (Crane et al., 2015), sedentary behavior reduction (Gardner et al., 2016) and promoting physical activities (Matthews et al., 2016). However, Credibility support strategies were the most strategies implemented in chronic arthritis apps (Geuens et al., 2016) whereas those strategies were absent in most reviewed mobile applications for physical activity (Matthews et al., 2016). Moreover, reminders were the most implemented strategy in medication management apps for consumers (Win et al., 2017). Our results revealed that self-monitoring, personalization, and reminder were the most frequently employed strategies in mental health apps. Some of these studies further reported more strategies that might not frequently be implemented in other health care apps (see **Table 2** and **Figure 3**).

TABLE 2 | Most strategies implemented in reviewed apps in other health domains.

References	Health domain	model used	Most strategies implemented
Crane et al. (2015)	Alcohol reduction	BCTs	• Self-recording (self-monitoring) • Information on consequences • Feedback on performance
Matthews et al. (2016)	Promoting physical activity	PSD	• Self-monitoring
Win et al. (2017)	Medication management	PSD	• Reminder • Tailoring • Self-monitoring
Gardner et al. (2016)	Sedentary behavior reduction	BCTs	• Setting behavioral goals • Social support • Instruction on how to perform the behavior • Self-monitoring
Geuens et al. (2016)	Chronic arthritis	BCTs and PSD	• Surface credibility • Expertise • General information • Self-monitoring
Our results	Mental health apps	BCTs and PSD	• Self-monitoring • Personalization • Reminder

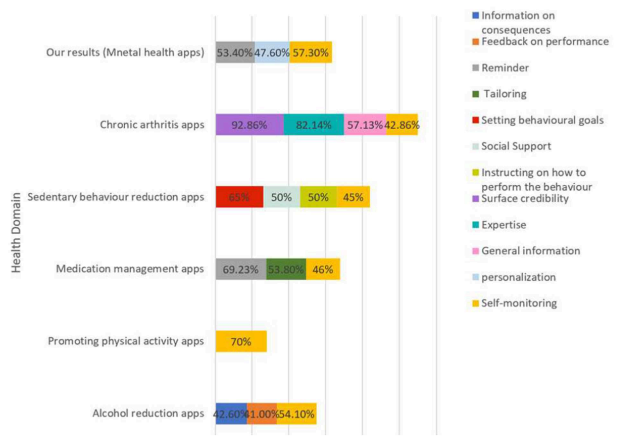


FIGURE 3 | Comparative the most strategies implemented in reviewed apps by other health domains.

Persuasive Strategies and Type of Mental Health Issues Targeted

We examined the persuasive strategies and the type of mental health issues the apps target. The results show that 65 apps target a combination of mental health issues whereas only 38 apps target a specific mental health issue. In general, the apps mostly targeted the following mental health issues: *anxiety, stress, depression,* and *general mental health* (see **Figure 4**). However, apps that targeted *stress* employed the highest number of persuasive strategies (23 out of 26 persuasive strategies identified in all mental health apps), followed by apps targeting *anxiety* and *depression* employed 20 persuasive strategies. **Figure 5** presents the overall number of persuasive strategies employed in each mental health issues. The results also show that personalization, self-monitoring and reminder were the most employed persuasive strategies in various mental health apps, see **Figure 3**.

Implementation of Persuasive Strategies

We present the various implementation of the common persuasive strategies in mental health application in this section.

Self-Monitoring Strategy

Self-monitoring strategy "allows people to track their own behaviors, providing information on both past and current" (Orji et al., 2017b, 2018c). In 23 apps, *self-monitoring* was implemented as being able to review trends of personal data related to mental health in a calendar or graphical format. Moreover, another 29

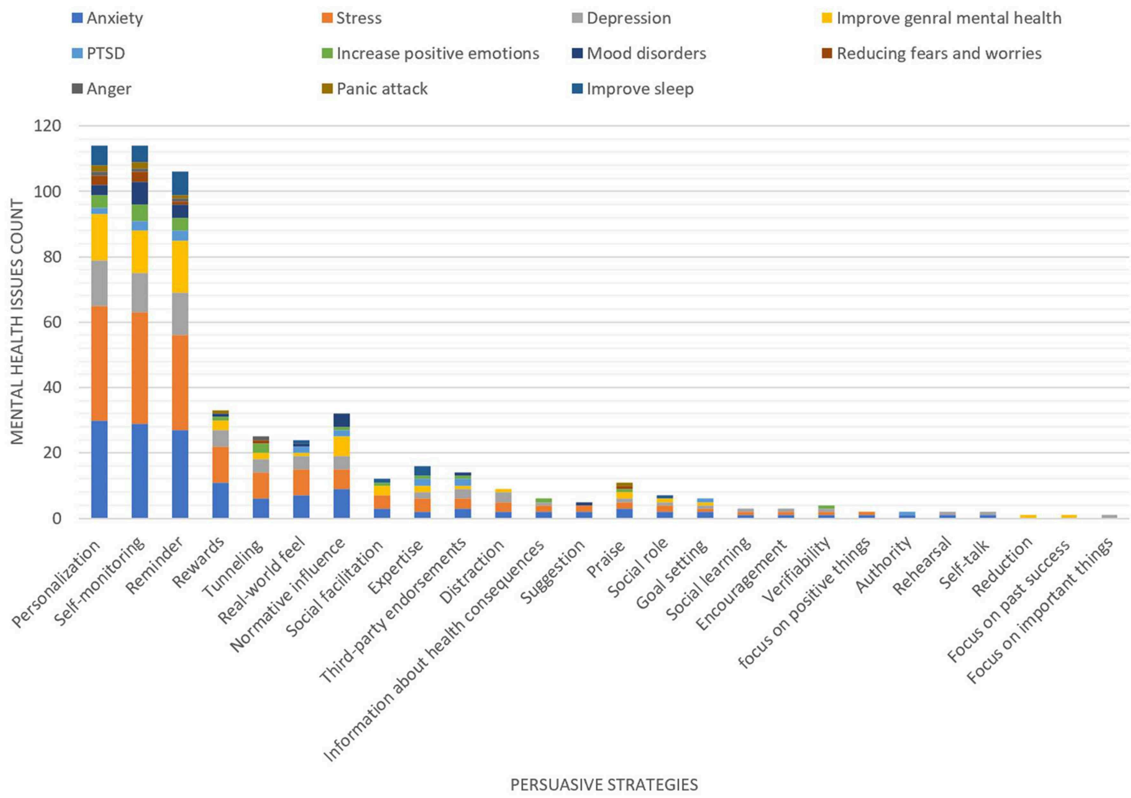


FIGURE 4 | Persuasive strategies for each mental health issue.

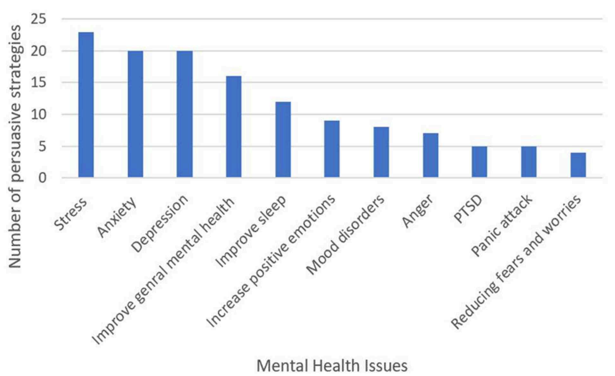


FIGURE 5 | Number of persuasive strategies employed in each mental health issues.

apps offer *self-monitoring* as a total number of activities related to mental health improvement performed by an individual and how long they spent on each activity. These include activities such as meditations.

Personalization

Personalization offers tailored contents, functionalities, and services to suit user's needs and choices. Tailoring content and functionality to a particular user's need based on his/her characteristics increases the efficacy of the system (Orji et al., 2017b, 2018c). Personalization was implemented in mental health apps in various ways including customizing the appearance of the app (such as background, theme, and sounds) to an individual's preference, customizing some app's functionalities (such as breathing rate, meditation duration, and music duration), providing some functionalities that allow individuals to adapt the apps to suit their personal preference (such as music, picture, activities, and challenges), tailoring the content based on certain user's characteristics (such as adapting meditation based on user's current emotion state).

Reminder Strategies

Reminder strategies enables a system to remind user to perform the target behavior. *Reminder* emerged as one of the popular strategies used in mental health apps. It is implemented in 49 mental health apps mainly to remind users to perform an activity (such as meditation, breathing, and assessment) or to track their personal data (e.g., mood). *Reminders* are often implemented as alert or pop-up boxes and sound.

Reward and Praise Strategies

Reward "offers virtual rewards to users for performing the target behavior" (Orji et al., 2017b) while **Praise** "applauds the user for performing the target behavior via words, images, symbols, or sounds as a way to give positive feedback to the user" (Orji et al., 2014). Mental health apps provide *reward* in various forms. Some mental health apps provide *reward* in form of points (6 apps), badges (8 apps), trophies (1 app), insight sticker (1 app), coins (1 app), planet growth (2 apps), and streak (1 app) that could be collected or gained while completing a task such as breathing and meditation. Only 2 apps implemented reward by allowing users to unlock more contents (such as more meditation sessions, or more lessons and activities) as a way of rewarding users. With respect to the Praise strategy, only 7 apps employed praise as words (i.e., Well Done) (6 apps) and colorful confetti (1 app).

Normative Influence Strategy

Normative influence strategy allows a system to provide "means for gathering together people who have the same goal and make them feel norms" (Harjumaa and Oinas-Kukkonen, 2009). *Normative influen*ce was implemented as a means for social interaction in form of community forums (13 apps) where users can exchange views about issues and feelings or by providing a link to join a Facebook community group (1 app).

Social Facilitation Strategy

Social facilitation strategy allows the system to provide a means for discerning other people who are performing the target behavior (Harjumaa and Oinas-Kukkonen, 2009). *Social facilitation* was implemented in the form of community forum of follower and following or listener and listening. Connected people can see each other's activities; followers can see the activities of the people they are following.

Credibility Strategy

Credibility strategy posits that systems should have competent look and feel (Harjumaa and Oinas-Kukkonen, 2009) to attract users and motivate behavior. The reviewed mental health apps employed many approaches to influence system credibility such as displaying **Expertise** (i.e., expertise in the design of the components of the app and knowledgeable in the information provided) (11 apps), providing **Real-world feel** [i.e., highlighting the people behind the design of the app (12 apps)], and by showcasing **Authority** [i.e., referencing notable organizations who are authority in the area of mental health (2 apps)]. Moreover, **Third-party endorsements** were employed in only 6 apps by providing a logo of respected sources such as a logo of a known university that is behind the app or approved the app. Moreover, few mental health apps implemented **Verifiability** (4 apps), allowing users to find more information by linking to studies or reports that provide evidence to support their claim or evidence that informed their design.

Tunneling Strategies

Tunneling strategies posits that system should guide users through the step-by-step process that lead to the target behavior by providing means for action that brings them closer to the target behavior. *Tunneling* was implemented in 13 mental health apps in the form of guidance on how to use the app to achieve a specific activity in line with the desired mental health outcome (e.g., how to meditate and how to breathe properly).

Encouragement

Encouragement was designed as a supportive message, and positive motivational quotes.

Focus on Positive Things

Focus on positive things was designed, as a game intended to help users learn how to focus on positive words or events as a way of promoting mental health and decreasing negativity.

Focus on Important Things

Focus on important things was implemented as a game intended to teach users how to focus on important things in users' life and avoid distraction and cognitive overload as a way of promoting mental health.

Distraction

Distraction was also implemented as a simple game aimed at diverting user's attention and distracting them from the current (negative) mood.

App Effectiveness and Persuasive Strategies Employed

To examine whether there is a relationship between the number of persuasive strategies employed in the app design and the perceived app effectiveness (as assessed by the app ratings), we performed Pearson's correlation between the app rating and the number of persuasive strategies. The results show that there is no relationship between the number of persuasive strategies and the app effectiveness; $r = 0.153$ (no correlation), $n = 103$, and $p = 0.123$. So, there is no significant correlation between the number of persuasive strategies and app rating, which demonstrates the perceived effectiveness from the user's point of view.

DISCUSSION

In this Section, we discuss our findings and offer some design recommendations for mental health app based on our findings.

Persuasive Strategies and Implementation

The purpose of this study is to identify distinct persuasive strategies incorporated in mobile apps designed to improve mental health and classify the persuasive strategies based on the type of mental health issues the app is focused on. Moreover, the study aims to reveal the various ways persuasive strategies are implemented/operationalized in mental health applications to achieve their intended objectives, and also to examine the relationship between apps effectiveness and the persuasive strategies employed.

Overall, the mental health apps reviewed in this paper employed 26 persuasive strategies, a range of 1 to 10 per app. However, 14% of the mental health apps did not employ any persuasive strategies.

Unsurprisingly, we found that self-monitoring is the most prevalent persuasive strategies implemented in mental health apps. According to Bakker et al. (2016), self-monitoring is considered the main feature of many evidence-based psychological therapeutic techniques such as cognitive behavior therapy, mindfulness exercises, emotion-focused therapy, Dialectical behavior therapy (DBT), and acceptance and commitment therapy (ACT). Self-monitoring help users with mental health issues to manage their conditions. They can track their feeling, thoughts, and behaviors which in turn increases self-awareness and improve mental health outcomes. The apps reviewed in this study limited tracking to manual input. Users with mental health issues need to enter their personal data manually which indeed is a major limitation. As highlighted by Orji et al. (2018a), manual recording is tedious, time consuming, and may not work for people with serious mental health issues.

Personalization emerged as the second most employed persuasive strategies in mental health apps. According to Price et al. (2016), personalizing some aspects of an app such as changing colors, setting backgrounds, and personalizing assessment questions would improve an app's usability. Most importantly, the ability to adjust the intervention delivered via mental health apps to suit the user's needs and characteristics will make the intervention more effective (Orji et al., 2014). Moreover, it has been found that personalized health interventions are more effective than the ones employing the one-size-fits all approach in other health domains (Orji, 2014; Orji et al., 2017a) and in depressive and anxiety disorders specifically (Carlbring et al., 2011; Silfvernagel et al., 2012). Therefore, the fact that most mental health app incorporated some form of personalization is not surprising considering that even people suffering from the same or similar mental health conditions may have unique needs that require individualized solutions.

Although tunneling and reduction reduces the effort required to achieve the target behavior by guiding the users to perform the task (tunneling) and by simplifying the task (reduction), it was incorporated only in 17 apps and 1 app, respectively. This is surprising, considering that individuals experiencing mental health conditions are often required to avoid stressful situations including complicated tasks that may stress them out and worsen their situation. Hence, making it essential that mental health apps are simple enough and also can guide users through the step-by-step process required to achieve the desired behavior is necessary to reduce the tendency of stressing them out when figuring things out themselves. Moreover, according to Alqahtani and Orji (2019), users with mental health issues complained about lack of guidance when using mental health apps which impair concentration and make them be easily frustrated. Therefore, reduction and tunneling strategy are essential for mental health apps.

The third persuasive strategy present in mental health applications is the reminder strategy. This strategy is mostly designed to remind users to track their personal data or to perform some mental health improvement activities such as meditation and breathing. Although reminders can help to increase adherence and reduce dropout from intervention, it was only implemented in 49 apps out of 103 apps. However, according to Bakker et al. (2016), a lot of annoying reminders can lead to disengagement. Therefore, developers should be careful when designing the reminder in mental health apps to avoid annoying people with frequent and unsolicited reminders. One way to achieve a balance between providing an effective reminder that will encourage users to adhere to the intervention and avoiding unnecessary reminders that will annoy user and make them disengage from the app is to tailor reminders to each individual. Individuals can be allowed to customize not only the frequency at which reminders are sent to them (how often), but also the type of reminder (pop up boxes, text message, sounds etc.) and when it should be sent (time).

Moreover, persuasive strategies such as reward and praise have been found to be among the popular strategies employed in many health apps to motivate users to be more engaged (Orji et al., 2014). However, only a few mental health apps implemented reward and praise, although apps for users with mental health issues may benefit from these techniques to motivate users. The reason why reward and praise are not as popular strategies in mental health apps is probably because many designers believe that improving mental health is an intrinsic reward of using their app hence no extra reward is required. Although, improving mental health is a major benefit and the main reason why many people would resort to using the app in the first place, it does not overwrite the need for extrinsic rewards such as badges and points, which have been shown to be effective at engaging users (Orji et al., 2013). According to Orji et al. (2018b), performing health behaviors is often difficult due to lack of immediate tangible benefit, offering intermediate rewards such as points and badges, may help to engage the users while the await the intrinsic reward.

The first credibility strategy is *real-world feel* and it is characterized by providing information about people or organization behind the app's content. This was found in 12 apps and still surprisingly low. In addition, 11 apps offered expertise in the design of the components of the app and information provided. We argue that these strategies are very important. All mental health apps, like all health apps, should provide information that is scientifically proven and evidence-based. Possessing the adequate technical skills to be able to develop an app is not enough for designing apps that will effectively improve or support mental health. Unfortunately, only 4 apps implemented the verifiability strategy which offers a way for users to verify the apps' content, 6 apps employed the third-party endorsements and 2 apps employed the authority which were very low. Credibility strategies are very important in mental health applications considering the sensitivity of the subject matter. Users need to be assured of not only the effectiveness and reliability of the app contents, but also that their data will be protected (privacy). In the Google store and App stores, anyone can design an app and publish it without providing evidence on the effectiveness of methods used in the app to manage the mental health issues and how users' information is protected. This results in a substantial number of apps that are not perceived as credible and trustworthy.

Social support is an important strategy for users who experience mental health issues because most of them often feel isolated or stigmatized. In this review, we found that only a few apps employed the strategies in the social support category. Only 17 apps employed the normative influence strategy which allow users to share their issues, thoughts, emotions with others to find support.

Certain strategies revealed in our study liken those identified in reviewed applications of other health domains. For example, self-monitoring strategies were similarly highlighted in reviewed applications of medication management apps for consumers (Win et al., 2017), chronic arthritis apps (Geuens et al., 2016), and promoting physical activities (Matthews et al., 2016). Moreover, self-monitoring strategies were specifically highlighted as a promising approach to sedentary behavior reduction (Gardner et al., 2016). However, it is worth mentioning that there are other strategies that emerge as most employed in the reviewed applications in other health domains that are not in mental health applications such as credibility support (Geuens et al., 2016).

In general, only a few persuasive strategies were employed in the apps that we reviewed which might explain the high attrition rate.

Persuasive Strategies and Type of Mental Health Issues

Most of the mental health apps that we reviewed targeted a combination of mental health issues which make it hard to know which persuasive strategies are more effective for a specific mental health issue. However, personalization, self-monitoring and reminder remain the most employed persuasive strategies in various mental health issues. Anxiety, stress, depression, and general mental health issues were the most issues the apps in this review target.

Apps Effectiveness and Persuasive Strategies Employed

The effectiveness of the app was measured based in the app's rating. Interestingly, we found no relationship between the number of persuasive strategies and apps effectiveness as indicated by users' ratings. This is particularly an interesting result considering the recent discussion and open research question on whether persuasive systems employing a multiple persuasive strategy are more effective than those employing a single strategy (Orji et al., 2017a). Our findings suggest that the number of strategies employed in apps design may not be related to the apps' effectiveness. According to Orji et al. (2017a), this is probably because employing a single appropriate strategy may be better than employing multiple inappropriate strategies or a combination of appropriate and inappropriate strategies that may have a cancellation effect. Hence, it is important that designers focus on selecting the appropriate persuasive strategies having both the target audience and the target behavior in mind.

Design Recommendation

Based on our findings, in this section, we offer some recommendations for designing mental health applications to improve users' adherence and engagement and hence apps' effectiveness. Moreover, some recommendations provided are from user app reviews, although the qualitative comments are not the focus of this work, we have integrated certain comments to support our recommendations.

1- Designer should employ self-monitoring in the apps that target mental health issues to help users to track their personal data and see their improvements over time. Allowing people with mental health issues to track and visualize their personal data in various format, would provide opportunity for self-awareness and help users take control of their mental health management. For example, "I love the app. It allows you to track emotions, experiences, discoveries, actions you took" [R29]. A major drawback is that most mental health app that employ self-monitoring use manual tracking which makes them tedious, time consuming, and users are likely to forget. To overcome this limitation, we suggest that designers employing self-monitoring should simplify the process and reduce the amount of work involved by automating behavior monitoring process whenever possible (Orji et al., 2018a) Although certain behavior data cannot be automatically tracked without users' involvements due to technology limitation. Therefore, for such behaviors that cannot be automatically monitored, designers should incentivize users, and reduce the perceived tediousness of the self-monitoring process using complementary persuasive strategies such as reminding users to log their behavior, rewarding users for tracking their behaviors each day, and reducing the number of steps required to record behavior" (Orji et al., 2018a).

2- Provide adaptive functionalities that allow users to adapt some app features such as the font size, font color, background, layout, type and length of meditation, breathing, and other mental health improvement tasks to suit each user's preferences and unique mental health needs. *For example, "The breathing exercises are great because I can set the type and time which I see as a great feature" [R487].* Personalization increases system relevance and usefulness (Orji et al., 2017c), enhances system's overall usability and ensure a personalized experience for each user. Moreover, adjusting app contents based on user's personal data will increase the effectiveness of the mental health interventions. In addition, since many mental health applications targets more than one mental health issues, it is necessary that the apps' content be adapted based on the type of mental health issues that users might be experiencing. However, even people suffering from the same or similar mental health conditions may have unique needs that require individualized solutions, hence highlighting the need to personalize mental health apps to each individual.

3- Provide an adequate reminder to remind user to track their data or to perform their meditation, breathing, and other mental health-related task. *For example, "I like a short notification "How are you feeling?" from time to time" [R213].* Although reminders can help to increase adherence and reduce dropout rate, a lot of annoying reminders can lead to disengagement. Therefore, developers should be careful when designing the reminder in mental health apps to avoid annoying people with frequent and unsolicited reminders. One way to achieve a balance between providing an effective reminder that will encourage users to adhere to the intervention and avoiding unnecessary reminders that will annoy user and make them disengage from the app is to tailor reminders to each individual. Individuals can be allowed to customize not only the frequency at which reminders are sent to them (how often), but also the type of reminder (pop up, text message, sounds etc.) and when it should be sent (time).

4- The mental health apps should provide a means for users to verify the reliability of their content and provide mental health information that are scientifically proven and endorsed by expert third parties. *For instance, "It works, and the science behind it is impressive" [R7]*. This will increase app credibility hence motivating users with mental health issues to engage with the intervention (Bakker et al., 2016). Possessing the adequate technical skills to be able to develop an app is not enough for designing apps that will effectively improve or support mental health. Credibility strategies are particularly important in mental health applications considering the sensitivity of the subject matter. Users need to be assured of not only the effectiveness and reliability of the app contents, but also that their data will be protected (privacy).

5- Mental health apps could benefit from implementing rewards and praise. For example, "*the growing tree is a nice way to see my practice is growing*" [R201]. Showing the growing tree as a kind of reward might motivate users to engage with the app. Moreover, providing users with mental health with motivational message when finishing the activities or task might encourage them continue using the app. Although designers may argue that improving mental health is an intrinsic reward of using their app hence no extra reward is required, however, it does not overwrite the need for extrinsic rewards such as badges, points, which has been shown to be effective at engaging users (Orji et al., 2013). According to Orji et al. (2018b), performing health behaviors is often difficult due to lack of immediate tangible benefit, offering intermediate rewards such as points, badges, may help engage the users while the await the intrinsic reward.

6- Employ Reduction and Tunneling to simplify mental health apps and guide users through the step-by-step process required to achieve the desired mental health outcome. *For example, "Users of the app are guided step by step in using every aspect to support their emotional health" [R73].* This will also reduce the tendency of stressing users out by allowing to figure things out themselves and hence reduce the overall dropout rate. Individuals experiencing mental health conditions are often advised to avoid stressful situations including complicated tasks that may stress them out and worsen their situation.

7- Employ the Social Support strategies in mental health apps (e.g., user forums) to provide users opportunity share their experience and support each other. Most people suffering from mental health issues often feel isolated or stigmatized, hence the need for social support. *For example, "It's a really good way to connect and feel connected to other people who have the same problem as you; even if you think you're alone" [R73].*

LIMITATIONS

This study has several limitations. Firstly, there exists the possibility that we missed some strategies due to the short timeframe of subsample applications. One method for overcoming this limitation is to extensively use a subsample for a longer duration to ensure no additional persuasive strategies are unrevealed. Secondly, user ratings are not enough to measure the effectiveness of apps because many other factors can affect the effectiveness of apps. However, user rating was the singular, closest evaluation we had to measure effectiveness.

CONCLUSION

In this paper, we deconstructed distinct persuasive strategies employed in 103 mental health applications using the Persuasive Systems Design (PSD) model and Behavior Change Techniques (BCTs). Two researchers independently coded 103 apps descriptions using the PSD model and BCTs. We further classified the persuasive strategies based on the type of mental health issues the apps aimed to address and how the strategies are implemented/operationalized in the mental health apps. The results show that self-monitoring, personalization, and reminder are the most commonly employed persuasive strategies in mental health apps irrespective of the mental health issues. We also found that anxiety, stress, depression, and general mental health are the mental health issues the apps mostly focused on. Above all, we uncovered that there is no relationship between the number of persuasive strategies employed and apps' effectiveness as measured using user ratings. We discuss various ways each persuasive strategy was implemented in mental health app to achieve the desired objective. Finally, we offered some design recommendations for mental health apps based on our findings. Future study should investigate which persuasive strategies are deemed as more important by users with mental health issues. We also hope to apply our recommendations in designing and evaluating mental health apps.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

FA designed the study. FA, OO, and GA collected the data. FA analyzed the data and wrote the manuscript. RO and OO reviewed the manuscript. RO supervised the study.

ACKNOWLEDGMENTS

We thank the NSERC Discover Grant for funding RO's research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2019.00030/full#supplementary-material

REFERENCES

Abraham, C., and Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health psychol.* 27, 379–387. doi: 10.1037/0278-6133.27.3.379

Almutari, N., and Orji, R. (2019). "How effective are social influence strategies in persuasive apps for promoting physical activity? A systematic review," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (Larnaca: ACM), 167–172.

Alqahtani, F., and Orji, R. (2019). "Usability issues in mental health applications," in *ACM UMAP 2019 Adjunct - Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization* (Larnaca), 343–348.

Arean, P. A., Hallgren, K. A., Jordan, J. T., Gazzaley, A., Atkins, D. C., Heagerty, P. J., et al. (2016). The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. *J. Med. Internet Res.* 18:e330. doi: 10.2196/jmir.6482

Bakker, D., Kazantzis, N., Rickwood, D., and Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Ment. Heal.* 3:e7. doi: 10.2196/mental.4984

Byrt, T., Bishop, J., and Carlin, J. B. (1993). Bias, prevalence and kappa. *J. Clin. Epidemiol.* 46, 423–429. doi: 10.1016/0895-4356(93)90018-V

Carlbring, P., Maurin, L., Törngren, C., Linna, E., Eriksson, T., Sparthan, E., et al. (2011). Individually-tailored, Internet-based treatment for anxiety disorders: a randomized controlled trial. *Behav. Res. Ther.* 49, 18–24. doi: 10.1016/j.brat.2010.10.002

Chang, T. R., Kaasinen, E., and Kaipainen, K. (2012). "Persuasive design in mobile applications for mental well-being: multidisciplinary expert review," in *International Conference on Wireless Mobile Communication and Healthcare* (Berlin; Heidelberg: Springer), 154–162.

Crane, D., Garnett, C., Brown, J., West, R., and Michie, S. (2015). Behavior change techniques in popular alcohol reduction apps: content analysis. *J. Med. Internet Res.* 17:e118. doi: 10.2196/jmir.4060

Fogg, B. J. (2009). "A behavior model for persuasive design," in *Proceedings of the 4th International Conference on Persuasive Technology* (Claremont, CA: ACM), 40.

Gardner, B., Smith, L., Lorencatto, F., Hamer, M., and Biddle, S. J. (2016). How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. *Health Psychol. Rev.* 10, 89–112. doi: 10.1080/17437199.2015.1082146

Geuens, J., Swinnen, T. W., Westhovens, R., de Vlam, K., Geurts, L., and Vanden Abeele, V. (2016). A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. *JMIR mHealth uHealth* 4:e118. doi: 10.2196/mhealth.6286

Harjumaa, M., and Oinas-Kukkonen, H. (2009). Persuasive systems design: key issues, process model, and system features. *Commun. Assoc. Inf. Syst.* 24: 485–500. doi: 10.17705/1CAIS.02428

Iacoviello, B. M., Steinerman, J. R., Klein, D. B., Silver, T. L., Berger, A. G., Luo, S. X., et al. (2017). Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. *JMIR mHealth uHealth* 5:e56. doi: 10.2196/mhealth.7226

Kelders, S. M., Kok, R. N., Ossebaard, H. C., and Van Gemert-Pijnen, J. E. (2012). Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J. Med. Internet Res.* 14:e152. doi: 10.2196/jmir.2104

Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *J. Consult. Clin. Psychol.* 73:539.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33:159. doi: 10.2307/2529310

Lehto, T., and Oinas-Kukkonen, H. (2011). Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. *J. Med. Internet Res.* 13:e46. doi: 10.2196/jmir.1559

Matthews, J., Win, K. T., Oinas-Kukkonen, H., and Freeman, M. (2016). Persuasive technology in mobile applications promoting physical activity: a systematic review. *J. Med. Syst.* 40:72. doi: 10.1007/s10916-015-0425-x

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., et al. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann. Behav. Med.* 46, 81–95. doi: 10.1007/s12160-013-9486-6

Orji, R., Lomotey, R., Oyibo, K., Orji, F., Blustein, J., and Shahid, S. (2018a). Tracking feels oppressive and 'punishy': Exploring the costs and benefits of self-monitoring for health and wellness. *Digit. Heal.* 4:2055207618797554. doi: 10.1177/2055207618797554

Orji, R., Mandryk, R. L., and Vassileva, J. (2017a). Improving the efficacy of games for change using personalization models, *ACM Trans. Comput. Interact.* 24:32. doi: 10.1145/3119929

Orji, R., Nacke, L. E., and Di Marco, C. (2017b). "Towards personality-driven persuasive health games and gamified systems," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO: ACM), 1015–1027.

Orji, R., Orji, F., Oyibo, K., and Ajah, I. A. (2018b). "Personalizing health theories in persuasive game interventions to gamer types: an African perspective," in *Proceedings of the Second African Conference for Human Computer Interaction: Thriving Communities* (Windhoek: ACM), 5.

Orji, R., Oyibo, K., and Tondello, G. F. (2017c). "A comparison of system-controlled and user-controlled personalization approaches," in *UMAP 2017 - Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (Bratislava), 413–418. doi: 10.1145/3099023.3099116

Orji, R., Tondello, G. F., and Nacke, L. E. (2018c). "Personalizing persuasive strategies in gameful systems to gamification user types," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC).

Orji, R., Vassileva, J., and Mandryk, R. L. (2013). LunchTime: a slow-casual game for long-term dietary behavior change. *Pers. Ubiquitous Comput.* 17, 1211–1221. doi: 10.1007/s00779-012-0590-6

Orji, R., Vassileva, J., and Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Model. User Adapt. Interact.* 24, 453–498. doi: 10.1007/s11257-014-9149-8

Orji, R. O. (2014). *Design for Behaviour Change: A Model-driven Approach for Tailoring Persuasive Technologies.* University of Saskatchewan. p. 1–257.

Oulasvirta, A., Rattenbury, T., Ma, L., and Raita, E. (2012). Habits make smartphone use more pervasive. *Pers. Ubiquit. Comput.* 16, 105–114. doi: 10.1007/s00779-011-0412-2

Price, M., Sawyer, T., Harris, M., and Skalka, C. (2016). Usability evaluation of a mobile monitoring system to assess symptoms after a traumatic injury: a mixed-methods study. *JMIR Ment. Heal.* 3:e3. doi: 10.2196/mental.5023

Roepke, A. M., Jaffee, S. R., Riffle, O. M., McGonigal, J., Broome, R., and Maxwell, B. (2015). Randomized controlled trial of superbetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. *Games Health J.* 4, 235–246. doi: 10.1089/g4h.2014.0046

Silfvernagel, K., Carlbring, P., Kabo, J., Edström, S., Eriksson, J., Månson, L., et al. (2012). Individually tailored internet-based treatment for young adults and adults with panic attacks: randomized controlled trial. *J. Med. Internet Res.* 14:e65. doi: 10.2196/jmir.1853

Torous, J., and Roberts, L. W. (2017). Needed innovation in digital health and smartphone applications for mental health: transparency and trust. *JAMA Psychiatry* 74, 437–438.

Wildeboer, G., Kelders, S. M., and van Gemert-Pijnen, J. E. (2016). The relationship between persuasive technology principles, adherence and effect of web-based interventions for mental health: a meta-analysis. *Int. J. Med. Inform.* 96, 71–85. doi: 10.1016/j.ijmedinf.2016.04.005

Win, K., Mullan, J., Katherine Howard, S., Oinas-Kukkonen, H., and Howard, S. (2017). "Persuasive systems design features in promoting medication management for consumers," in *Hawaii International Conference on System Sciences*.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 20 March 2020
doi: 10.3389/frai.2020.00009

Toward a Taxonomy for Adaptive Data Visualization in Analytics Applications

Tristan Poetzsch[1], Panagiotis Germanakos[2] and Lynn Huestegge[1]**

[1] Department of Psychology, Julius-Maximilians-University Würzburg, Würzburg, Germany, [2] User Experience ICD, Product Engineering, Intelligent Enterprise Group, SAP SE, Walldorf, Germany

Data analytics as a field is currently at a crucial point in its development, as a commoditization takes place in the context of increasing amounts of data, more user diversity, and automated analysis solutions, the latter potentially eliminating the need for expert analysts. A central hypothesis of the present paper is that data visualizations should be adapted to both the user and the context. This idea was initially addressed in Study 1, which demonstrated substantial interindividual variability among a group of experts when freely choosing an option to visualize data sets. To lay the theoretical groundwork for a systematic, taxonomic approach, a user model combining user traits, states, strategies, and actions was proposed and further evaluated empirically in Studies 2 and 3. The results implied that for adapting to user traits, statistical expertise is a relevant dimension that should be considered. Additionally, for adapting to user states different user intentions such as monitoring and analysis should be accounted for. These results were used to develop a taxonomy which adapts visualization recommendations to these (and other) factors. A preliminary attempt to validate the taxonomy in Study 4 tested its visualization recommendations with a group of experts. While the corresponding results were somewhat ambiguous overall, some aspects nevertheless supported the claim that a user-adaptive data visualization approach based on the principles outlined in the taxonomy can indeed be useful. While the present approach to user adaptivity is still in its infancy and should be extended (e.g., by testing more participants), the general approach appears to be very promising.

Keywords: graph adaptivity, data visualization, user model, analytics, graph ergonomics, recommendation engine

OPEN ACCESS

Edited by:
Alexandra Ioana Cristea,
Durham University, United Kingdom

Reviewed by:
Lei Shi,
Durham University, United Kingdom
Armando Maciel Toda,
Universidade de São Paulo São Carlos, Brazil

****Correspondence:***
Tristan Poetzsch
poetzsch.tristan@gmail.com
Lynn Huestegge
lynn.huestegge@uni-wuerzburg.de

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change,
a section of the journal Frontiers in Artificial Intelligence

Received: *13 October 2019*
Accepted: *21 February 2020*
Published: *20 March 2020*

Citation:
Poetzsch T, Germanakos P and Huestegge L (2020) Toward a Taxonomy for Adaptive Data Visualization in Analytics Applications. Front. Artif. Intell. 3:9. doi: 10.3389/frai.2020.00009

INTRODUCTION

As the recent acquisition of analytics application provider Tableau by software giant Salesforce shows, the relevance of self-service data visualization software is rapidly increasing. Considering the associated commoditization, the user group for data analytics applications is not only becoming larger, but also more diverse, and so are personal backgrounds and levels of experience regarding data visualizations (Convertino and Echenique, 2017; Lennerholt et al., 2018). Although dealing with diversification is thus becoming more relevant, current research still focuses either on data-based recommendations for visualization generation (Viegas et al., 2007; Vartak et al., 2015; Wongsuphasawat et al., 2016) or on individual factors determining the processing of data visualizations (e.g., the data literacy concept) (Gal, 2002; Shah and Hoeffner, 2002; Roberts et al., 2013). However, these two important areas have not yet been sufficiently considered in conjunction, although the benefits of improving the accessibility of data through individualized visualizations may very tangibly contribute to achieving better business decisions. Hence, in this paper we explore the potential of the latter idea by enriching the current user models with specific human characteristics, and by following an experimental approach we propose a taxonomy for user adaptivity in data visualization as a foundation for further research.

In order to derive this taxonomic approach, we outlined three main research questions for this paper: Is there a need for a user-adaptive approach to data visualization? How should a taxonomy be structured in its user-adaptive approach? Can the usefulness of an adaptive approach be validated? Based on these research questions we structured the present paper, starting with an examination of previous works on data visualization. Based on this, we determined a user model consisting of user traits, states, and strategies as well as their respective operationalizations.

Following up on the theory, we conducted three studies to understand the need and the relevant factors for an adaptive approach to data visualization. *Study 1* explored how User Interface Design experts would visualize different data sets, thereby addressing the first research question regarding the need for an individualized approach. We hypothesized that recommendations by the experts would vary significantly, therefore supporting the need for an individualized approach. For the second research question on how a taxonomy should be structured, we conducted two follow-up studies based on the proposed user model. *Study 2* explored how user *traits* impact on the perception of different data visualization encodings, and hence laid the groundwork for adapting to traits. The associated hypothesis was that not all visualizations were suitable for every user. *Study 3* focused on understanding how user *states* can be operationalized as intents, and how these can differ from each other. Here the hypothesis was that different intents are characterized by different associated cognitive subtasks and should therefore significantly impact on visualization requirements. Both studies are necessary preconditions contributing to the design of an adaptive data visualization taxonomy.

Based on these insights, a general adaptive taxonomy of diagram choice, layout, and specific visualization design was derived. An important feature of this taxonomy is that it can handle multidimensional data and state- or trait-related user variables. The validation for the usefulness of the taxonomy as outlined in the third research question was addressed in *Study 4*, in which User Interface Design experts were asked to perform different tasks with visualizations suggested by the taxonomy and rated their experience afterwards. The findings indicated some potential for such a taxonomy, although there is still some work to be done before it may be applied in a consumer setting.

BACKGROUND AND PREVIOUS WORK

Data visualizations have been in use to present numerical information since the early twentieth century (Eells, 1926) and consequently spawned a research tradition that is still active (Cleveland and McGill, 1984; Shneiderman, 1992; Heer et al., 2010). Four larger research streams can be summarized under this umbrella. While at first the focus was on optimizing single visualizations (Gillian and Lewis, 1994), the requirement to display more complex information subsequently led to research into how multiple charts may be layouted next to each other in a "multi-view" perspective (Roberts, 2007). The rise of personal computers then enabled not only static displays, but also the possibilities to interact with data visualization, supported by the field of human factors (Stasko et al., 2008). Lastly, the research on automatic recommendations for data visualizations became increasingly relevant as more users without an academic degree in statistics or computer science gained access to self-service analytics solutions (Mackinlay, 1986; Stolte and Hanrahan, 2000; Wongsuphasawat et al., 2016). All these research streams combine challenges from various fields, including psychology, computer science, and human factors.

Before exploring these research streams in more detail, some terms need to be disambiguated first. A chart refers to a single visualization of a set of data, for example, a bar chart. Within such charts, an indicator refers to a graphical element that represents the value of a single data point associated with a variable, such as a single bar in a bar chart or a point in a scatterplot. The type of encoding/indicator refers to the kind of indicator in use, such as bars, points, or lines (Gillian and Lewis, 1994). Finally, another important distinction for the present purpose is that between adaptivity, personalization, and customization. While adaptivity refers to a system that automatically sets up functionality and a user interface to fit the user, personalization requires the user to actively set up the system in a way that fits him/her. In contrast, customization takes place when a third party sets up the system to fit the user (Germanakos, 2016). The following literature review on data visualization will refer to this basic terminology.

Research on data visualization was first concerned with the optimization of single chart visualizations, starting with Eells (1926). Corresponding research on data encoding effectiveness peaked in the 80s and 90s, when landmark studies like those by Cleveland and McGill (1984), who invented dot plots, or by Hollands and Spence (1992), who evaluated line charts vs. bar charts as the most effective means to communicate change in data (see also Huestegge and Philipp, 2011; Riechelmann and Huestegge, 2018), emerged. Scatterplots, on the other hand, were later considered an optimal choice for visualizing correlations (Harrison et al., 2014; Kay and Heer, 2016). Over the years, new visualization techniques such as tree maps were introduced (Shneiderman, 1992; Heer et al., 2010; Bostock et al., 2011). Besides encoding types, particular features of visualizations like color (Lewandowsky and Spence, 1989; Demiralp et al., 2014) or chart size were studied more closely. Regarding the latter, several studies emphasized that smaller charts ($<17^\circ$ of the visual field) were considered helpful in avoiding gaze shifts along with associated inaccuracies (Heer et al., 2009; Heer and Bostock, 2010; Strasburger et al., 2011; Orlov et al., 2016). Apart from data point reading accuracy, the size of a chart was also shown to influence perceptual strategies: While smaller graphs facilitated quick overall assessments and immediate responses to graphs, larger charts led to increased scrutiny during graph comprehension (Orlov et al., 2016).

As most data sets and real-world contexts are too complex to be displayed in a single chart, more research on multi-view visualizations emerged in the early 2000's. These flexible data visualization features also became more prominent in data analytics software. For example, pioneering applications such as snap-together visualization (North and Shneiderman, 2000) or Polaris (Stolte and Hanrahan, 2000) emerged. Historically, such multi-view visualizations originally consisted of dual-views of data (Roberts, 2007), comprising, for example, Overview + Detail, Focus + Context, or Difference Views, the latter involving two datasets that are laid out next to each other to facilitate comparison. Another line of research on more complex data sets focused on so-called "Small Multiples," which depict the relationship of several variables relative to each other. This type of visualization was also combined with a master-slave approach, so that manipulating data in one view also affected visualizations in the other view (Roberts, 2007; Scherr, 2009; van den Elzen and van Wijk, 2013).

The development of these types of data visualization also stimulated research on interaction with corresponding graphs. Especially as interactive visualizations became more and more common at the end of the twentieth century, studies on using interactive graphs, which quickly became a standard in data analysis software, were on the rise. This was not surprising, since interactive visualizations offer many benefits for working with data, from providing context information to increasing attention (Stasko et al., 2008). With the onset of touch-based devices, an entirely new class of interactive data display solutions, with its own set of challenges, emerged: Especially with dashboards, main goals for designing applications for mobile devices comprised maximizing the size of each visualization, minimizing occlusion, keeping all visualizations in view, and reducing any need for end-user customization of views (Sadana and Stasko, 2016).

Finally, with increasing commercial interest in data visualization for large sets of data, automation of data visualization became an important issue. Self-organizing dashboards based on recommendation systems were developed as an answer to the disproportionally large amount of user time devoted to data handling (compared with the actual goal of conducting science; Howe and Cole, 2010). Automatic data visualization recommendations have come a long way (Mackinlay, 1986; Stolte and Hanrahan, 2000; Viegas et al., 2007; Vartak et al., 2015; Wongsuphasawat et al., 2016). Especially due to the commoditization of data analytics, recent recommendation engines such as Voyager 2 (Wongsuphasawat et al., 2017) are gaining increasing attention. The underlying criteria for these recommendation systems are best outlined along the axes data characteristics, intended task or insight, semantics and domain knowledge, visual ease of understanding as well as user preferences and competencies (Vartak et al., 2017). Taken together, research in this area already points toward further development of recommendation systems in the areas context sensitivity and, ultimately, user-adaptive data visualization.

STRUCTURING INDIVIDUAL PERCEPTION OF DATA VISUALIZATION WITHIN A USER MODEL

As mentioned above, adapting the display of data to the user is the next challenge in the field of data visualization, especially since information overload appears to become a major problem in business decisions (Moore, 2017), and therefore calls for user-specific approaches. However, structuring user-adaptive data visualization requires a user model in the first place (Germanakos, 2016). A useful basic distinction in this context is that between (relatively persistent) user traits and (more transient) situational states of the user (Kelava and Schermelleh-Engel, 2008). Based on various possible traits and the states, several strategies can be applied by users to deal with visualized data, as displayed in **Figure 1**.

Research on how individual traits can affect the perception of data visualization started around the 80's with the concept of graphical literacy, loosely defined as "*the ability to read and write (or draw) graphs*" (Fry, 1981, p. 383). Later, the concept was elaborated, and subdivided into the three skill levels "*reading the data,*" "*reading between the data,*" and "*reading beyond the data*" (Friel et al., 2001; Okan et al., 2012). Based on these theoretical considerations, fostering the development of graphical literacy became a focus of research (Gal, 2002; Shah and Hoeffner, 2002; Roberts et al., 2013). Furthermore, determining the cognitive variables underlying graphical literacy has also been of considerable interest. Commonly, perceptual speed, visual working memory and—to some extent—verbal working memory were discussed as potentially relevant factors in this regard, sometimes joined by locus of control (Velez et al., 2005; Conati and Maclaren, 2008; Toker et al., 2013; Lallé et al., 2015). Perceptual speed refers to the "*speed in comparing figures or symbols, scanning to find figures or symbols, or carrying out other very simple tasks involving visual perception*" (Conati and Maclaren, 2008, p. 202). Both verbal and visual working memory are part of the working memory architecture proposed by Baddeley (1992). Specifically, the visuospatial sketchpad comprises the ability to manipulate visual images, while the phonological loop stores and rehearses speech-based information. As perceptual speed and visual working memory have repeatedly been shown to be relevant traits regarding the perception of data visualization, we chose to include assessments of these constructs in our studies.

When outlining user states and user intentions in particular within the data analytics context, no widely accepted general model is currently present in the literature. It has been proposed that, on a higher processing level, one should distinguish visualization purposes into analysis, monitoring, planning, and communicating (Few, 2004). However, it should be noted that specifying the actual intention of a user and providing the appropriate information is certainly not a trivial challenge (Gotz and Wen, 2009; Conati et al., 2015; Oscar et al., 2017).

Previous research on strategies for reading data visualizations primarily focused on visual processing, reasoning with data points, and integrating context knowledge (Amar et al., 2005;

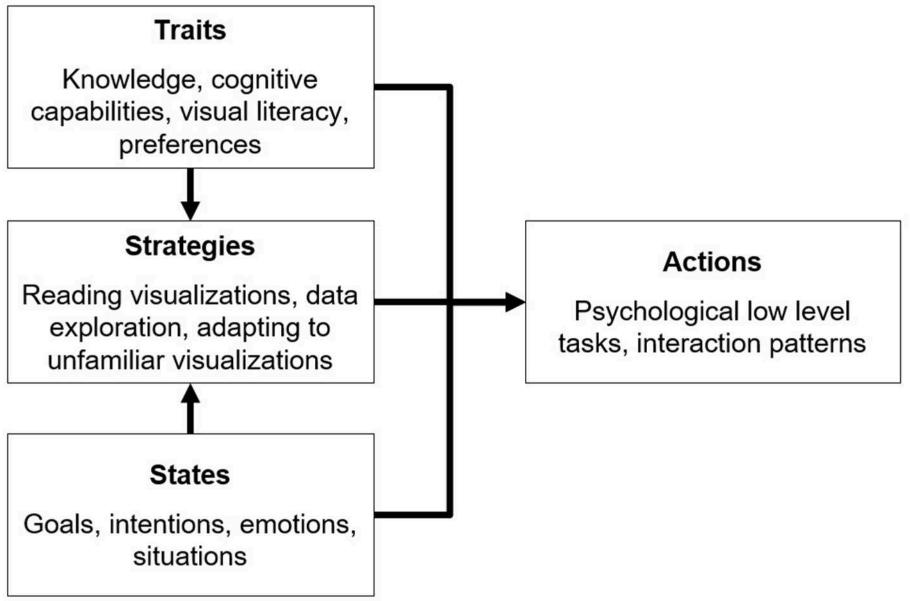


FIGURE 1 | Modeling aspects of user-adaptive data visualization.

Ratwani et al., 2008). Interindividual differences in strategies were assumed to be especially relevant in more unrestricted settings such as understanding unfamiliar visualizations (Lee et al., 2016), or in the context of designing visualizations (Grammel et al., 2010).

EXPLORING THE NEED AND THE RELEVANT FACTORS FOR THE DESIGN OF AN ADAPTIVE TAXONOMIC APPROACH

Although previous literature repeatedly recommended to put more research effort into studying user adaptivity in the context of data visualization (see above), we reasoned that it is mandatory to verify the need for an underlying taxonomic approach to user state/trait-based visualizations first. Therefore, we conducted *Study 1* with User Interface Design experts ($n = 16$) to explore the need for an adaptive taxonomic approach by letting them freely design visualizations for several data sets in order to see if the results would differ, thereby underlining the need for adaptivity.

Based on this, we conducted two follow-up studies to understand user traits and states in more detail based on a user model. *Study 2* ($n = 45$) had the goal of evaluating how individual traits and backgrounds affect the interpretation of different data visualization types. Therefore, the number of errors in the interpretation as well as interpretation speed were measured with 20 different visualizations. *Study 3* evaluated how different goal-states affected the decision which cognitive processing steps are taken in working with data visualization, and how—based on the taxonomy—visualizations should therefore adapt based on user goals. Two main goal-states (analysis vs. monitoring) were identified from theory, and analytics experts regularly working with numbers ($n = 14$) were questioned about their typical real-life tasks involving data. These tasks were split into their associated (low-level) cognitive processing steps and classified into one of the two goals. Based on the results of the conducted studies, a taxonomy was derived (see section Developing a User-Adaptive Visualization Taxonomy). All data sets from the studies are available online (link in section **Supplementary Materials**).

Study 1—Exploring the Need for an Adaptive Approach

Study 1 explored how User Interface Design experts would visualize different data sets. As we already anticipated the need for an adaptive approach based on an underlying taxonomy, we hypothesized that analytics expert participants would vary considerably in choosing a type of encoding for various given data sets.

Participants

Sixteen analytics experts (SAP employees, 7 female, 9 male) were tested and interviewed. The participants were chosen based on their experience within the context of data analytics. All participants had an academic background and were working for SAP for at least 6 months in the areas of UX or analytics. All participants also had considerable experience in working with data visualization (>1 year of professional experience). The age range was 27–48 years.

Stimuli

There were 16 data sets to be visualized by the participants. These data sets were of different complexity and constructed based on the combinations of four dichotomous factors: (a) single or multiple (three) numerical variables, (b) single or multiple (three) categorical variables, (c) data including or excluding time as a variable, and (d) data with 1:1 cardinality or a 1:n cardinality. In six of these data sets, geographical variables were included.

Procedure

After a brief introduction, the data sets were presented to the participants one after another on single sheets of paper by the experimenter. The participants were asked to sketch recommendations for respective visualizations on the same page as the data set. This was done in order to minimize a potential influence of software restrictions or software experience. The encoding recommendations were classified by the experimenter. The study lasted around 42 minutes ($SD = 13$).

Design and Data Analysis

The independent variables for designing the data sets were the number of categorical variables (1 or 3) in the dataset, the number of numerical variables (1 or 3) in the dataset, the cardinality of the data (1:1 or 1:n) and if a time variable was presented (yes or no). There was one data set for each combination of these independent variables. For each data set, we calculated the proportion of different (vs. same) visualizations designed across participants (e.g., a proportion of 100% would indicate that all participants came up with the same solution), which served as the dependent variable.

Results

The solutions proposed by the participants varied considerably, as indicated by a mean proportion of different visualizations of 51% (SD = 20.4). In some cases, all participants proposed different types of data visualizations. A one-sample *t*-test indicated that the mean significantly differed from 100%, $t_{(15)} = 9.97$, $p < 0.0001$. Therefore, the results generally support the assumption of substantial variability in individual visualization preferences. In addition, it was observed that most participants actively tried to reduce data complexity by plotting multiple, differently scaled numerical variables on the same axis (sometimes even stacking these differently scaled variables, resulting in misleading data representations). Another strategy to reduce complexity was the use of filters. Finally, we also observed that intra-individual consistency in chart choice (e.g., always using a geomap for geodata) tended to be low. As an additional exploratory analysis, we also conducted a multiple regression analysis using the independent variables (for designing the data sets) as predictors. This analysis resulted in a significant overall regression ($R = 0.91$, $p < 0.001$) with significant contributions of the predictors "number of categorical variables" ($\beta = 0.66$, $t = 5.276$, $p < 0.001$) and "number of numerical variables" ($\beta = 0.58$, $t = 4.620$, $p = 0.001$), while the remaining two predictors had no significant impact ($\beta < 0.21$, $t < 1.7$, $p > 0.12$). Specifically, an increase in the number of variables led to more diverse solutions.

Discussion

The variability present in the results of this exploratory study generally supports the call for user-adaptive data visualization. Participants suggested different visualizations for the same data sets, even though general user characteristics such as their academic background and field of work were relatively similar. In the context of the proposed high-level user model, the results therefore suggest that it may be worthwhile to study potential effects of more specific user traits and strategies on visualization selection and design to eventually optimize and support visualization decisions. In addition, as most participants had problems with dealing with the inherent complexity of the data, a taxonomic approach that not only takes user variables but also (multidimensional) data characteristics into account would clearly be desirable.

Limitations

There were several shortcomings in this exploratory study that need to be discussed. First, due to time restrictions eight of the 16 participants were not able to complete all tasks, and thus the results of the analyses should only be interpreted with great care. Second, the approach of this study lacked some degree of ecological validity, as participants were asked to choose visualizations without the help of a dedicated software. In a brief interview at the end of the study protocol, several participants commented that they would actually click through all available alternatives in a given software instead of actively developing a visualization concept. Third, all participants were employees of SAP and therefore almost certainly affected by the company's design language and typical visualization solutions, even though the heterogeneity of the results implied that this did clearly not result in similar outcomes among participants. However, one might suspect that the results might vary even more substantially if analysts or analytics UI experts from other companies were added to the sample.

Study 2—Examining the Perception of Various Visualizations Considering User Traits

As *Study 1* suggested the need for a user-adaptive approach to data visualization, the next step was to derive more specific research questions based on the user model outlined above. Starting with user traits, *Study 2* aimed at becoming more specific about determining which traits may be relevant for the development of an adaptive taxonomy, especially regarding the selection of specific types of visual encoding. Based on results from previous literature (see above), we specifically focused on prior experience, visual literacy, and cognitive capacities. These factors were considered relevant for the participants' ability to understand and work with a wide variety of data visualizations. More specifically, we hoped that it is possible to classify individual data visualizations into those more suitable for experts or novices in order to take this issue into account within the taxonomy. This was done using a cluster analysis approach.

Consequently, the main hypothesis in this study was that some visualizations are more appropriate for participants with substantial prior experience, visual literacy, or cognitive capacities to adapt quickly to these visualizations. Prior experience was operationalized in terms of education, working in a data-driven job, and the degree of statistical knowledge. To measure graphical literacy, the *Subjective Graphical Literacy Scale* (SGL) (Galesic and Garcia-Retamero, 2011; Garcia-Retamero et al., 2016) was used. Cognitive capacities were tested by assessing perceptual speed and visual working memory. To measure perceptual speed, a *Sum to 10 test* (Ackerman and Beier, 2007) was used as it is also based on numerical (and not only visual) abilities. In this test, participants are presented with combinations of two numbers, and have to decide quickly if their sum is equal to 10 or not. To measure visual working memory, a *Visual Patterns Test* based on the Visual Patterns Test by Della Sala et al. (1999) was administered. In this test, participants are exposed to a black and white pattern grid and have to recognize this pattern from a selection of similar patterns after a brief distraction interval. A multiple regression analysis was used to test which of these several traits significantly predict graph comprehension abilities.

Participants

All participants ($N = 45$) were recruited via Social Media. Thirty-three participants were male (73%), 12 Participants were female (27%). Thirteen of the participants had a high school degree, 19 a bachelor's degree and 12 a master's degree, and one had finished an apprenticeship.

Stimuli

The following data visualization types were evaluated: Scatterplot, Area Chart, Stacked Bar Chart, Stacked Area Chart, Boxplot, Bullet Chart, Waterfall Chart, Bubble Chart, Heatmap, Treemap, Sunburst, Sankey Chart, Matrix Scatterplot, Trellis Bar Chart/Small Multiples, Sparklines and Horizon Charts (see Heer et al., 2010, for details on these visualization types).

Procedure

The experiment was web-based and therefore completed on the participants' own devices. The experiment was designed and conducted using the platform soscisurvey (www.soscisurvey.com). After a brief introduction to the study and its parts, participants were asked to report their highest educational degree. After that, they were asked if their job involved a lot of work with numbers and graphs on the scale "No"—"Sometimes"—"Yes." Additionally, participants were asked to rate their familiarity with statistics and data interpretation on the following scale: "Not familiar at all"—"Somewhat familiar (e.g., familiar with averaging)"—"Familiar (e.g., familiar with correlation, variability measures, different types of distributions including normal distributions)—"Very familiar (e.g., familiar with factor analysis, cluster analysis, ANOVA)." After this, participants completed the SGL, the Sum up to 10 test, and the Visual Patterns Test. In the main part of the study, participants were provided with 16 data visualization types (see above). For all diagrams, participants were asked how familiar they were with this diagram and how good they think they could handle this diagram on a scale from 1 ("Not at all") to 5 ("Very good"). After that, participants were provided with 4 statements about the data, which they had to judge as either correct or incorrect (including the option to choose "I don't know"). The study lasted about 19 ($SD = 3{,}5$) min.

Design and Data Analysis

Graphical literacy, performance in the visual working memory (VWM) test (number of errors and response time), the performance in the perceptual speed (PS) test (number of errors and response time), and prior experience (education, working in a job where charts and numbers are common and statistical knowledge) served as independent variables in the analyses. The dependent variables comprised the participants' general understanding of the graph (choosing "I don't know" instead of answering the questions), the number of errors made on the tasks with different visualizations, and task solving time. For analyses, an exploratory cluster analysis was conducted, followed by multiple linear regressions (see below for details).

Results

First, to gain an understanding of how the tested visualizations may be related to each other regarding their general susceptibility to errors as well as the self-ratings with respect to the general understanding of these visualizations, a hierarchical cluster analysis was conducted based on the two dependent variables "general understanding" and "number of errors," which were considered most relevant from an applied perspective. The resulting dendrogram/clusters are shown in **Figures 2**, **3**. Three clusters were derived in total. The first cluster contained error-prone visualizations. The second cluster comprised multivariate and hierarchical visualizations, which were associated with more errors, and which were partially difficult to understand. The third cluster combined all sub-optimal visualizations. These visualizations are common, but do not support error-free interpretation. Horizon charts represented an outlier among all visualizations, as it was both difficult to understand and frequently misinterpreted. Thus, this type of visualization should therefore be generally avoided. The main hypothesis was that some visualizations are more appropriate for participants with substantial prior experience, visual literacy, or cognitive prerequisites necessary to quickly adapt to these visualizations. To address this main hypothesis, all visualization types that were not understood by all participants were grouped into an "expert cluster" (consisting of bullet charts, boxplots, matrix scatterplots, sankeys, and bubble charts). For these visualizations, multiple regression analyses were conducted in order to model the impact of independent user variables on the general understanding, the error rates, and the reaction times as stated in the hypothesis. An overview of results for an initial multiple regression analysis based on general understanding is shown in **Table 1**. However, note that 18 observations had to be excluded due to missing data on two independent variables (participants who were not active in a job), and the regression model was only marginally significant, $F_{(9,17)} = 2.48$, $p = 0.051$, adj. $R^2 = 0.339$. A second model only considered the factors statistical knowledge and VPT errors (which were the only significant predictors within the full model) and resulted in a significant effect overall, $F_{(2,42)} = 6.41$, $p < 0.01$, adj. $R^2 = 0.197$, including all cases. Finally, as visual working memory is probably difficult to measure in an applied software setting, a third model focusing on statistical knowledge only was calculated. This analysis also resulted in a significant prediction of the general understanding of data visualizations, $F_{(1,43)} = 6.93$, $p < 0.02$, adj. $R^2 = 0.119$. A correlation analysis verified the analysis, as statistical education $r_{(43)} = -0.37$, $p < 0.02$ and visual working memory $r_{(43)} = 0.36$, $p < 0.02$

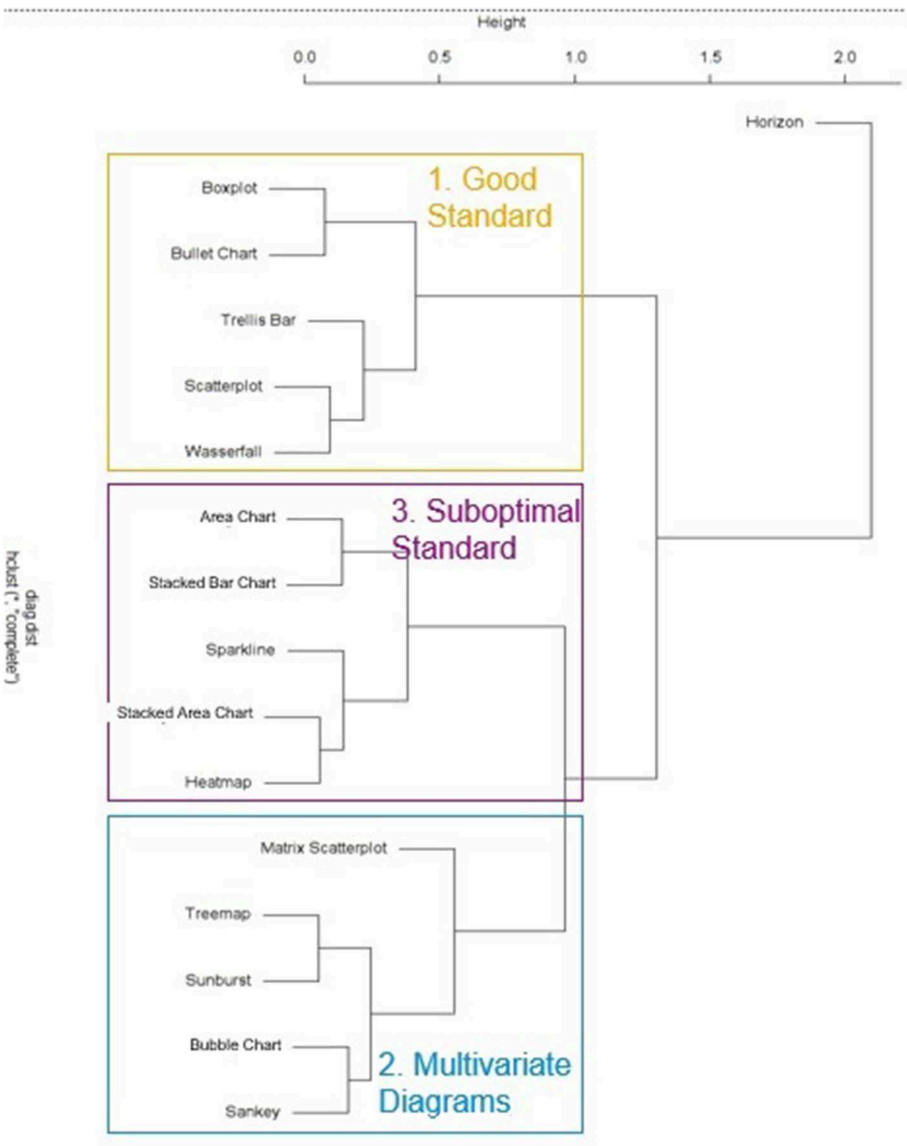


FIGURE 2 | Study 2: Dendrogram of visualization clusters.

were significantly correlated with each other. Multiple linear regressions were also run for analyzing error rates, as shown in **Table 2**. In an initial analysis using all predictors, the factor education was significant, but the whole model was not, $F_{(9,17)} = 1.04$, $p = 0.450$, adj. $R^2 = 0.013$. This model was based on the sample with 19 missing cases on two independent variables (participants who were not active in a job). When excluding the factors responsible for the missing cases (job experience) in a second model, no factor was significant anymore and the model again was not significant, $F_{(7,34)} = 0.55$, $p = 0.790$, adj. $R^2 = -0.083$. Finally, only using the factor education as a predictor also yielded no significant effect, $F_{(1,43)} = 1.03$, $p = 0.315$, adj. $R^2 = -0.001$. The self-reported familiarity with the provided expert data visualization types significantly predicted the number of errors made in the tasks, $F_{(1,43)} = 4.17$, $p < 0.05$, adj. $R^2 = 0.067$, although the effect is not particularly strong.

Discussion

This study identified three clusters of visualization types, namely the "good standard" (Scatterplot, Trellis Chart, Waterfall, Boxplot and Bullet Chart), the "suboptimal standard" (Sparkline, Heatmap, Stacked Bar Chart, Area Chart, Stacked Area Chart), and the "multivariate visualizations" (Sunburst, Treemap, Matrix Scatterplot, Sankey Chart, Bubble Chart). An additional (artificially created) cluster combined Boxplot, Bullet chart, Sankey Chart, Bubble Chart and Matrix Scatterplots into an "expert visualizations" group. Understanding of these visualizations was predicted significantly by the users' statistical

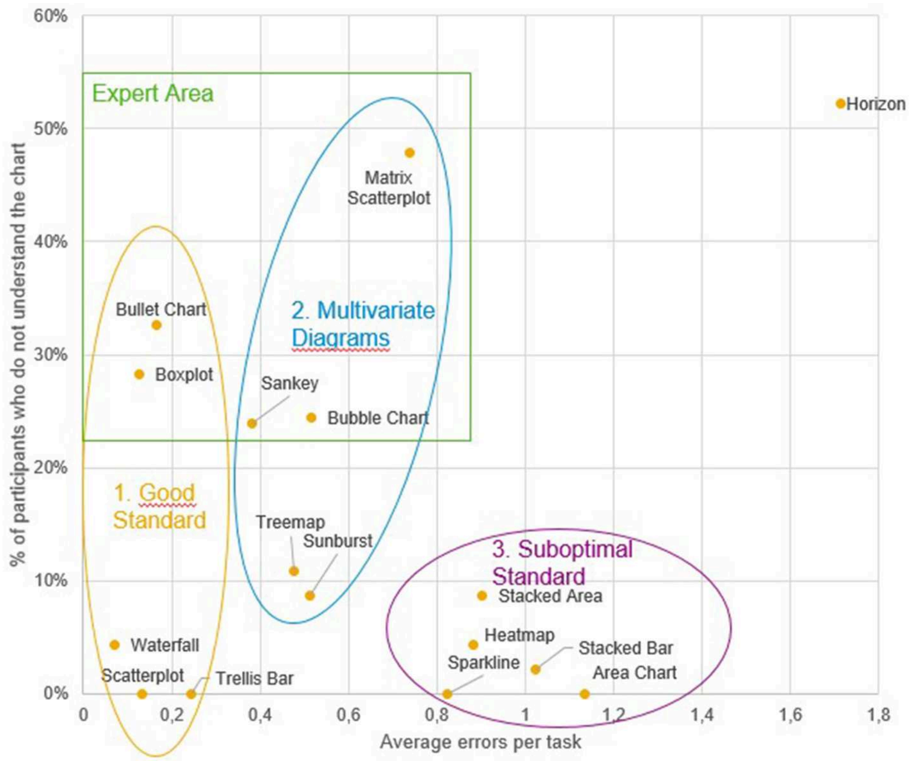


FIGURE 3 | Study 2: Scatterplot with visual representation of the emerging clusters and the artificial expert cluster.

TABLE 1 | Study 2: multiple regression models to predict the general understanding of visualizations in the expert cluster.

Factor	Variable	*B*	*SE B*	*β*	*t*	*p*
MODEL 1 (18 MISSING)						
Graphical literacy	SGL	0.01	0.06	0.03	0.19	0.856
Cognitive variables	VWM error	0.69	0.28	0.48	2.46	0.025*
	VWM time	0.01	0.02	0.07	0.36	0.726
	PS errors	−0.22	1.37	−0.03	−0.16	0.875
	PS time	−0.10	1.02	−0.18	−0.01	0.922
Prior experience	Education	−0.25	0.22	−0.23	−1.16	0.261
	Job – Numbers	0.10	0.44	0.05	0.23	0.818
	Job – Graphs	−0.39	0.40	−0.23	−0.98	0.341
	Statistical knowledge	−0.83	0.38	−0.50	−2.16	0.045*
MODEL 2						
Cognitive variables	VWM error	0.52	0.23	0.31	−2.28	0.028*
Prior experience	Statistical knowledge	−0.65	0.27	0.33	−2.42	0.020*
MODEL 3						
Prior experience	Statistical knowledge	−0.73	0.28	−0.373	−2.632	0.012*

*$p < 0.05$.

knowledge, and therefore this should be considered a crucial factor in providing recommendations for a user. Interestingly, neither the self-reported ability to work with charts nor the SGL score were good predictors for either the probability of understanding a chart or the errors made in the tasks. Also, despite a current debate emphasizing the impact of cognitive variables on learning to work with data visualizations (Velez et al., 2005; Conati and Maclaren, 2008; Toker et al., 2013; Lallé et al., 2015), the effect of perceptual speed on the ability to adapt to unfamiliar data visualizations (or to work more accurately with them) could not be replicated in this study. Visual working memory, on the other hand, was indeed a significant predictor of the ability to understand unfamiliar visualizations (but not of error-free reasoning with them).

Limitations

This study also suffered from several limitations. One major limitation was that participants only had to complete a single task based on each visualization type. As the difficulty of the tasks was not controlled independently, the evaluation of visualization types may be quite vulnerable to task-based processing disruptions or task difficulty. Additionally, participants were not a representative sample, as the educational background was nearly exclusively academic. A more diverse sample may yield more nuanced results and provide answers to the question of the extent to which people with lower educational levels can work with different types of visualizations. Furthermore, the online setting of the present study is not a controlled environment and therefore potentially subject to distraction. This may have also influenced the measurement of cognitive abilities, although it may also be argued that in a work setting distractions can actually be considered to occur quite frequently.

TABLE 2 | Study 2: multiple regression models to predict the errors in the expert cluster.

Factor	Variable	B	SE B	β	t	p
MODEL 1 (18 MISSING)						
Graphical literacy	SGL	−0.01	0.02	−0.10	−0.46	0.653
Cognitive variables	VWM error	0.03	0.09	0.07	0.31	0.759
	VWM time	0.00	0.01	0.17	0.72	0.481
	PS errors	−0.51	0.42	−0.30	−1.23	0.235
	PS time	−0.51	0.31	−0.38	−1.64	0.119
Prior experience	Education	−0.15	0.06	−0.53	−2.23	0.039 *
	Job – Numbers	0.00	0.13	−0.00	0.00	0.999
	Job – Graphs	0.15	0.12	0.35	1.24	0.233
	Statistical knowledge	0.09	0.16	0.21	0.75	0.464
MODEL 2						
Graphical literacy	SGL	−0.02	0.02	−0.16	−0.97	0.339
Cognitive variables	VWM error	−0.00	0.07	−0.01	−0.07	0.946
	VWM time	0.00	0.00	0.01	0.06	0.954
	PS errors	−0.37	0.33	−0.20	−1.15	0.258
	PS time	−0.22	0.22	−0.18	−0.99	0.329
Prior experience	Education	−0.04	0.04	−0.16	−0.93	0.358
	Statistical knowledge	0.02	0.07	0.04	−0.227	0.822
MODEL 3						
Prior experience	Education	−0.04	0.04	−0.15	−1.02	0.315

*$p < 0.05$.

Study 3—Examining the Relationship of User Intents and Low-Level Actions

After examining a selection of user traits in Study 2, the next step was to focus on another aspect closely associated with user traits, namely the more transient user states (e.g., emotions, intentions etc.). Based on the four dissociable user intents monitoring, analyzing, planning, and communicating (assumed to be engaged in a cyclical fashion, see Few, 2004), the two intents monitoring and analyzing were selected for Study 3. We reasoned that these two intents were more closely associated with perceiving and understanding data visualizations, while planning and communicating were rather related to deriving actions. In order to distinguish between analyzing and monitoring, we first focused on low-level task profiles. Specifically, several analysts and managers were interviewed regarding their regular work with data visualizations and the associated tasks. For low-level tasks, we distinguished between the sub-tasks retrieving values, filtering, computing a derived value, finding extrema, sorting, determining ranges, characterizing distributions, finding anomalies, clustering, and correlating (Amar et al., 2005). It was hypothesized that the user intents "analysis" and "monitoring" are associated with significantly different patterns of these low-level sub-tasks. If this holds true, the corresponding visualizations should therefore be different, too.

Participants

Fourteen experts (4 female, 10 male) were interviewed. They were recruited via personal network and were questioned via telephone. The participants were from different departments in different companies, ranging from sales management in a small e-commerce startup to controlling in a DAX-30 automotive corporation. All participants had an academic background, and the age range was 25–55.

Procedure

After a short introduction and some information regarding the background of the study, all participants were asked to report which data-related tasks (involving visualizations) they were frequently engaged in. One participant could principally report any number of user tasks. After the interviews, each reported user task was assigned to either a monitoring or an analysis intent, and then they were further decomposed into their low-level sub-task components (see above). The reported user tasks were assigned to a monitoring intent if they comprised a check against a point or level of comparison and produced a binary result (e.g., "Control if work hours in every department indicate overtime"). Otherwise, they were assigned to the analyzing intent (e.g., "Checking how much plan and actual were apart in last periods of time").

Results

The 14 participants reported 45 tasks altogether (average 3.2 tasks per participant). Of these 45 tasks, 19 were analysis tasks and 26 were monitoring tasks. By calculating a t-test for the mean number of low-level tasks associated with each intent, a significant difference could be observed, $t_{(43)} = 5.397, p < 0.0001$. Specifically, tasks associated with an analysis (vs. monitoring) intent involved a significantly greater number of low-level sub-tasks per reported task. A chi-square test also revealed that the distribution of the occurrence of the 10 sub-tasks involved in the two different intent types significantly differed, $\chi^2 (9) = 42.61, p < 0.001$. Therefore, our hypothesis was confirmed. The distribution of basic tasks within the two types of intents is illustrated in **Figure 4**.

Discussion

The hypothesis of this study was that the low-level task profiles of monitoring and analysis user states are significantly different. Based on the results of the study this hypothesis can be confirmed, therefore using these intents as a basis for adapting visualizations to user states appears to be reasonable. Because of the high level of relevance for all intents, the retrieval of values as well as computing derived values should in particular be as easy as possible. A major implication of this study was that highlighting anomalies in a monitoring setting is an important feature in data visualizations. Being able to highlight specific aspects, however, implies reserving one (ideally pre-attentively processed) feature specification (e.g., color) for callouts. Following up on this thought, reserving colors for semantic callouts may be considered advisable for a monitoring setup. This would need visualizations to be charted without colors by default.

Limitations

This study also had some shortcomings. A major problem was that both the classification of reported user tasks to intents and low-level tasks to reported user tasks was essentially subjective. However, we believe that our criteria were overall quite reasonable, and it was necessary to start at some point.

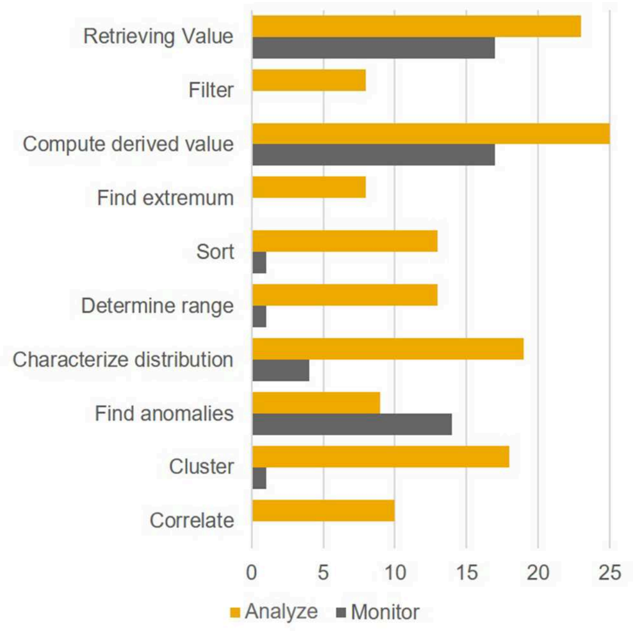


FIGURE 4 | Study 3: Distribution of low-level tasks within the intents.

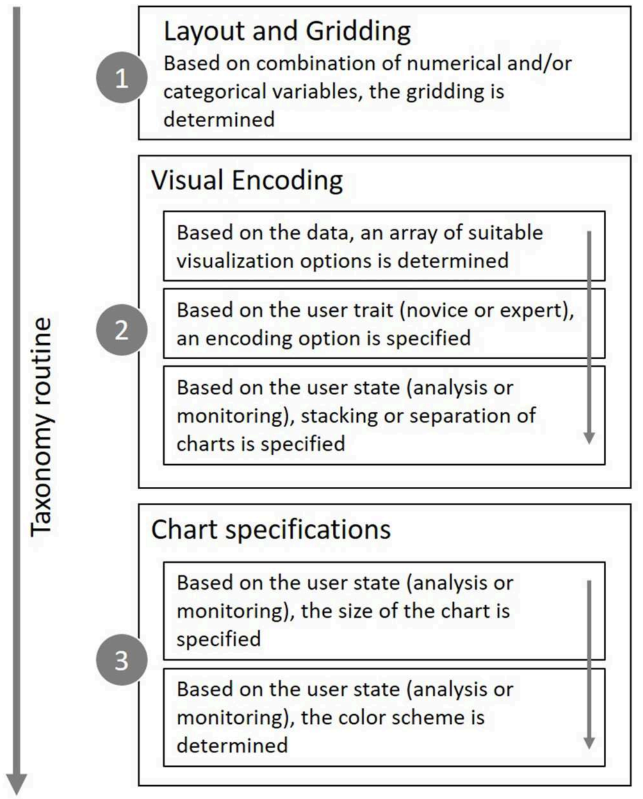


FIGURE 5 | Adaptive taxonomy calculcation steps to determine layout, encoding and specifications.

Nevertheless, a more objective classification would be desirable. Also, the sample size was relatively small, and once again only academic participants were assessed. The resulting implications discussed above therefore may thus not be generalizable to user groups with lower education levels.

DEVELOPING A USER-ADAPTIVE VISUALIZATION TAXONOMY

The main aim of this paper was to provide a first approach for developing systematic user-adaptive visualization recommendations. While previous related studies (see above) can be used as clear guidelines for the general design of data visualizations, the studies described so far represent a reasonable basis for our decision to include adaptive elements based on the user trait "expertise" (Study 2) and the user state "intention" (Study 3). To make the taxonomy applicable for all kinds of data sets, a central requirement was also to provide a structured approach for the visualization of both simple and complex variable settings. The proposed solution, which is mainly grounded on discussions with experts and own prior experience with typical options present in modern data visualization software, is outlined in the following section and summarized in **Figure 5**.

Layout and Gridding of Data

If a given data set has more dimensions than can (or should) be displayed in a single chart, a layout of several charts is needed. This grid should combine categorical variables on one axis and numerical dimensions on the other axis (C-N-Matrix). In this way, n-dimensional data sets can be visualized. If there are only up to two numerical variables, these can be displayed directly in the chart, and the second grid axis may also be used for categorical variables (C-Matrix), which is the equivalent of a pivot table. Both the C-N-Matrix and the C-Matrix are displayed in **Figure 6**. While the C-N-Matrix is the most flexible way to visualize any data set, the C-Matrix may be more space efficient and should therefore be used preferentially.

Visual Encoding of Queries

Choosing a specific visual encoding type defines which kind of indicators are used to encode the actual data values. For data sets with a 1:1 relationship, bar charts are usually recommended as a visual encoding type in both monitoring and analysis settings due to consistent reports of their superiority over alternative encoding types (e.g., Cleveland and McGill, 1984; Heer and Bostock, 2010; Huestegge and Poetzsch, 2018). However, if one of the variables represent time, a line chart is usually recommended as it promotes the mental processing of developments over time. When a 1:n relationship is present in the data, the default for the monitoring intent should be to aggregate the data (e.g., averaging) and to display it as a bar chart, although there should be an easily accessible control element for switching back to the raw data. Given an analysis intent, a boxplot should be the first choice to display the data distribution in a condensed way, if possible enriched with a violin to account for distribution nuances (Matejka and Fitzmaurice, 2017). If the user lacks

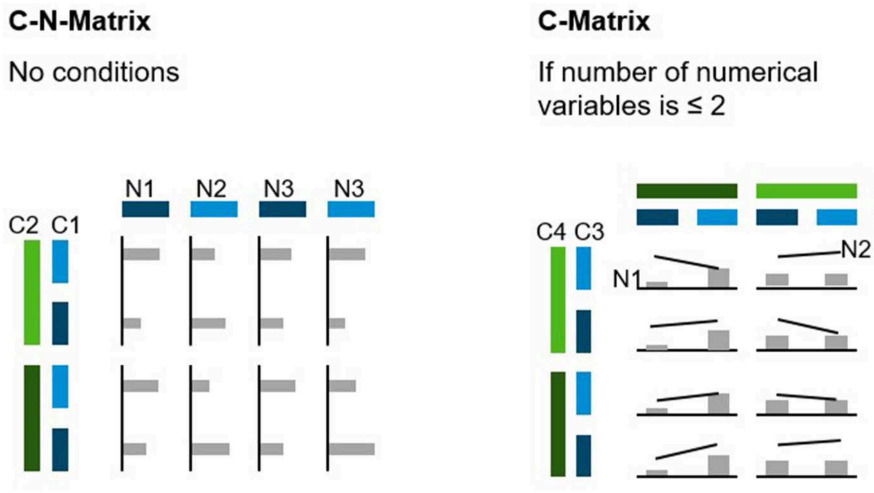


FIGURE 6 | C-N-Matrix and C-Matrix.

sufficient statistical background knowledge or the rendering engine is not able to display this type of graph, the second-best option would be a distribution curve. When distribution visualizations are not available at all, there is no other option than to display the raw data points. For this, strip plots should be preferred over normal dot plots, as the former allow for a much tighter packing of indicators without a substantial risk of "over-plotting."

It is principally possible to shift one variable from the categorical grid axis to the data encoding region itself through stacking of indicators. The most frequently encountered example for this option is the stacked bar chart. Stacking comes with the benefit of offering the possibility to focus on combined values for comparison purposes (e.g., comparing spending categories across departments). However, beside this distinct benefit stacking may also decrease the speed and accuracy at which a user judges trends within a category, mainly due to the more complicated cognitive demand of aligning and judging indicators without a common baseline (Simkin and Hastie, 1987), as also shown in *Study 2*. Following the general outline of the monitoring intent, which mainly focuses on getting a quick overview over a complex data pattern, a separate display of charts seems to be a reasonable default, while stacking seems to be useful for an analytic intent when the variable at hand represents a sum (not an average, as summing averages is not useful in most contexts).

Deriving Chart Specification Recommendations

Although the value encoding type is the most prominent feature of any visualization, specifications such as the size of a chart and its coloring can also affect the perception and understanding of charts. Thus, these features should also be considered, in particular as a function of user intent. The optimal chart size in the context of a monitoring intent certainly cannot be determined exactly. However, it should be large enough to allow for an accurate, readable depiction of the visual indicators, but at the same time as small as possible to prevent unnecessary shifts of visual attention (Heer et al., 2009; Heer and Bostock, 2010; Orlov et al., 2016). Such an optimal size should also entail that a numerical scale should roughly fit in the foveal area (5° of the visual field), or at least the parafoveal area (about 8°). Additionally, chart sizing should be flexible enough to account for multiple devices. A useful unit of measurement in this context may be the root em (rem), which is usually considered a standard size in current web design. While desktop setups and devices with lower resolution convert 1 rem to 16 pixels (px) during rendering, high-resolution devices usually transform 1 rem to 32 px (Powers, 2012). This is supposed to ensure optimal readability, as the x-height is above the 0.2° threshold (Legge and Bigelow, 2011). Modeling the optimal rem size for different device scenarios across the visual field (Kaiser, 1996), a sizing of 10 rem has been considered to represent a good choice. If the data are separated, the individual charts can be decreased in size down to around 5 rem, which should still result in accurately readable charts. In the context of an analysis intent, it may be beneficial to provide a larger chart, as this presumably facilitates a more thorough and specific exploration of the data (Orlov et al., 2016). However, the size of the chart should not exceed perifoveal vision, as a significant decrease in stimulu detection occurs beyond 20° of the visual field (see section "Related Previous Works"). The radius of perifoveal vision in the example described above is equivalent to about 333 px. The largest fitting rectangle would be a square with a length of 471 px or 29 rem. Nevertheless, it remains to be considered that the width of the chart depends on the number of data points/categories at hand, and therefore the actual width of a chart may well-exceed the recommended size. For Scatterplots and other encoding types involving numerical variables on both dimensions, these size recommendations apply for both dimensions. Regarding the coloring of charts in a monitoring setting, it may be considered beneficial to refrain from using colors or to restrict coloring to a few desaturated indicator colors (Few, 2009).

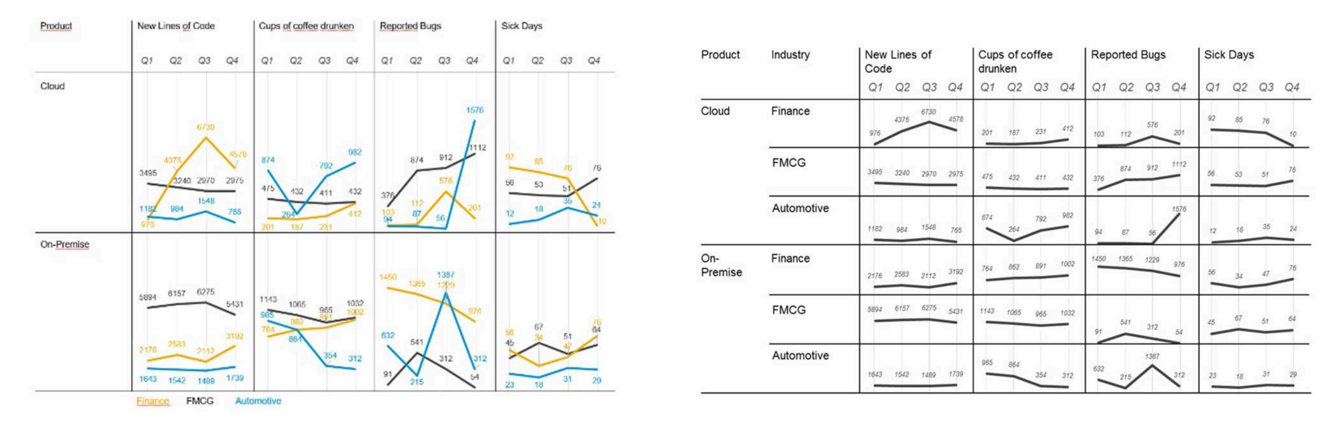


FIGURE 7 | Two visualizations displaying the same data set for an analysis setting **(left)** and a monitoring setting **(right)**.

As a separation of charts is proposed for this type of setting, the colors are likely not needed immediately and can instead be reserved for semantic callouts (e.g., warnings) to increase the visibility of such callouts. For analytic settings, especially those involving stacked variables, more colors may be needed, although desaturated colors may reduce mental distraction in these cases, too.

STUDY 4—VALIDATING THE TAXONOMY WITH EXPERTS

The taxonomy outlined in the previous section certainly needs empirical validation to prove its usefulness for data visualization recommendation systems. As a first step into this direction, *Study 4* therefore evaluated whether data visualization experts considered the taxonomy-based recommendations for different settings suitable in the context of tasks that closely resemble real-world applications. The hypotheses evaluated in this study were as follows: The visualizations provided by the taxonomy are generally judged as suitable by the experts. Furthermore, we tested whether the taxonomy is suited to visualize even complex data sets without the resulting charts being judged as significantly less suitable than in simple settings.

Participants

Ten analytics experts from within SAP were interviewed. All participants were male. Seven of the 10 participants held at least a master's degree. The participants reported to frequently work with data, with an average self-rating of 4.4 on a scale from 1 "never" to 5 "very often" ($SD = 0.84$). They judged their statistical education level to be at an average of 3.2 on a scale ranging from 1 to 4 (see Study 1 for details on this scale). The age range was 25–51.

Stimuli

The participants worked through 12 trials, each consisting of a task and an associated visualization. These 12 trials were built from six data sets, which were each combined with both a monitoring task and an analysis task (in separate trials). The specific visualizations varied across task types as suggested by the taxonomy. The six data sets were characterized by three different degrees of complexities: The easiest settings consisted of three dimensions (one categorical, one numerical, and one time variable), the intermediate settings consisted of five dimensions (two categorical, two numerical, and one time variable), and the most complex settings involved seven dimensions (two categorical, four numerical, and one time variable). Example stimuli are shown in **Figure 7**.

Procedure

The experiment was paper-based and completed in presence of an experimenter. After a brief introduction, demographic data were gathered, including age, education, and how much the participants worked with data in their job. Additionally, participants were asked to report their statistical education. In the main part of the experiment, the participants were provided with 12 trials (six monitoring tasks and six analysis tasks). In the monitoring tasks, participants were asked to "mark the data point(s) or dimension(s) you found to stick out and may need deeper analysis." In the analysis settings, participants were provided with four statements about the data set in a multiple-choice format and asked to mark the correct options. After completing the task in each trial, participants were asked to judge how suitable the given visualization was for the task on a scale from 1 (not suitable) to 5 (very suitable). Each trial ended with an open section for the participants to provide feedback and optimization ideas. All data sets were taken from the field of enterprise performance management, which the participants could relate to. The average study time was 40 min ($SD = 7.55$).

Design and Data Analysis

Trial setting (analysis/monitoring) and data set complexity (3/5/7 variables) served as independent variables. The dependent variables were the suitability rating scores. For hypothesis testing, a within-subject two-way ANOVA was conducted.

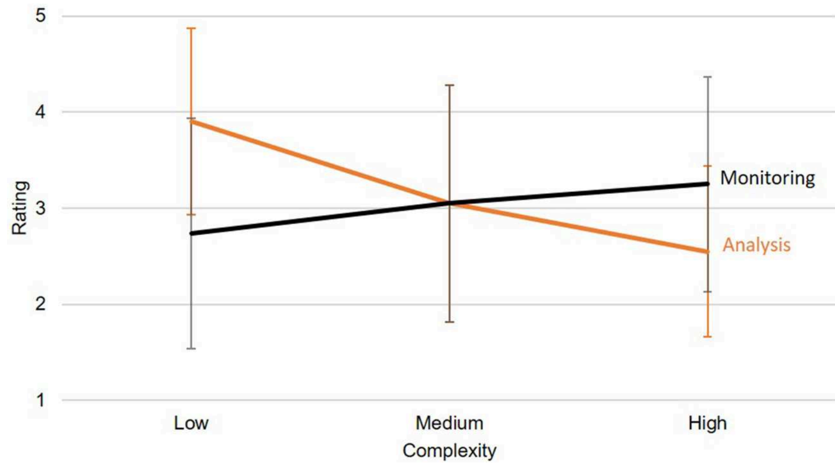


FIGURE 8 | The ANOVA showed a significant interaction effect between the task and the complexity regarding the rating. Error bars show standard deviations.

Results

Results regarding the first hypothesis, namely whether the visualizations are generally judged as suitable by the experts, were rather ambivalent. The average rating amounted to 3.09 ($SD = 1.17$) on a scale from 1 to 5, suggesting that the visualizations proposed based on the taxonomy were rated at an intermediate level of suitability. However, these numbers only reflect one part of the picture: In the verbal feedback for each visualization, it became clear that a lot of factors negatively affected the ratings that were not part of the main study rationale: Unit measures were reported to be unfitting, variable choices were considered invalid regarding their contextual validity, and features such as filtering and aggregating were missed as they were not possible on paper. After the experiment, all participants were familiarized with the idea of an adaptive data visualization taxonomy, and all of them endorsed the idea. Therefore, although the quantitative data are not finally conclusive in this regard, the verbal feedback supported the general usefulness of a taxonomic approach. The second hypothesis proposed that the taxonomy is also useful in visualizing even complex data sets without being judged as significantly less suitable than for simple data sets. To answer this question, a two-way ANOVA was calculated for the experts' ratings. There was no significant main effect for either data complexity, $F_{(2,18)} = 1.06$, $p = 0.369$, nor for task, $F_{(1,9)} = 0.55$, $p = 0.476$. However, we observed a significant interaction, $F_{(2,18)} = 11.47$, $p < 0.001$, $\eta p^2 = 0.560$. While monitoring settings with higher complexity were rated better than those with lower complexity, this effect was reversed for analysis settings, as shown in **Figure 8**. It may be argued that in analysis settings, complex data are usually not processed based on a single view, but rather experienced sequentially by interacting with the data using filtering, brushing, aggregating, and drilldown options (which were not available in the present study).

Discussion

The main aim of this experiment was to provide first evidence for the usefulness of the developed taxonomy for the adaptive display of data. Although it could not be concluded that the specific recommendations derived from the taxonomy were already sufficiently suitable for an instant implementation into existing applications, the general principles of the taxonomy were embraced by our sample of experts. Based on the verbal reports, it appeared that some aspects were already sufficiently useful, for example the scalability in the monitoring mode for scenarios with different complexity. In contrast, the display options for the analytic mode were less well-suited, therefore more work in this regard is necessary.

Limitations

A main limitation of this study is that user traits were not considered here as the study was conducted in a paper-based manner. Therefore, a fully user-adaptive approach could not be evaluated here. Also, some side effects of our decision to use a paper version negatively affected the suitability ratings (see above). Regarding the stimuli used in this study, it would have been useful to compare different visualizations for each intent/task combination. Through this procedure, it may have become more evident where benefits or disadvantages of specific visualizations are located, and if the taxonomy actually provided good recommendations when compared with other possible solutions. This approach appears to be promising for further research, which should also involve larger sample sizes.

CONCLUSION AND FUTURE WORK

A central claim of the present paper is that data visualizations should be adapted to both the user and the context. This idea was supported by *Study 1*, which demonstrated substantial inter-individual variability among a group of experts when freely choosing an option to visualize data sets. To lay the theoretical groundwork for the envisioned taxonomic approach, a user model combining user traits, states, strategies, and actions was proposed and further evaluated empirically in *Studies 2* and *3*. The results implied that for adapting to user traits, statistical expertise is a relevant dimension that should be considered. Additionally, for adapting to user states different user intentions such as monitoring and analysis should be differentiated and accounted for. These results were used to develop a taxonomy which adapts visualization recommendations to these (and other) factors. For example, a monitoring intention may benefit from separated data lines without coloring, while an analysis intention should benefit from combined charts. In addition to this adaptive approach, the taxonomy also outlined a way to grid up complex data sets to optimize their visualization. A preliminary attempt to validate the taxonomy in *Study 4* tested its visualization recommendations with a group of experts. While the corresponding results were somewhat ambiguous overall, some aspects of the results nevertheless supported the claim that a user-adaptive data visualization approach based on the principles outlined in the taxonomy can be useful. Of course, the present approach to user adaptivity is still quite rudimentary, especially due to the relatively low number of participants. To solidify the results, larger samples should be collected.

In theory, one might want to consider every user to be an individual based on multiple (potentially quantitative) dimensions relevant to visualization adaptivity. Here, we only considered very few of these dimensions, which usually comprised two binary alternatives (e.g., two task-based intents). Thus, more research is needed in order to finally come up with a truly individualized output. Although adaptivity in the context of data visualization is still in its initial stage, it clearly has a lot of potential for future development. The full potential of adaptive visualization will probably be of great relevance especially in more complex settings of decision support involving data visualization, where tailoring information width and depth to the user is mandatory.

One of the areas that need to be worked on more extensively in order to move forward with adaptive visualizations is the contextual component, as "*to create useful adaptive visualization tools we must understand the relationship between a users' context and the visualization they require*" (Oscar et al., 2017, p. 811). Without knowing more about the context, it is only possible to provide a sensible default. One possible option to address this issue could be the use of conversational user interfaces, which would allow the user to articulate context and intentions in more detail and subsequently enable the system to provide more suitable visualizations. Another approach to adaptivity would be to let the system collect data about usage patterns and then to suggest these learned patterns to users later. This approach, also known as collaborative filtering, requires large amounts of data and is essentially theory-blind: While it does not need any theoretical assumptions in order to work, it cannot take into account basic knowledge about which behavioral, user- or context-related aspects can be meaningfully combined. Due to this serious problem, it may be concluded that knowledge-based filtering may represent a reasonable middle way (Vartak et al., 2017).

Another area crucial to the implementation of such a taxonomy the issue of standardization. This is relevant for both the design of charts as well as for how visualizations are coded. For example, chart rendering engines are usually not compatible to each other. A single system is not able to control different engines as their required input format differs, although first steps toward a standardized encoding format have been taken with the introduction of *CompassQL* (Wongsuphasawat et al., 2016). Nevertheless, the corresponding problems do not only affect rendering engines: A micro-service-based approach to the whole data visualization ecosystem could also encompass data query recommendation or statistical modules. This would provide a truly flexible system that could not only adapt visualizations, but also adapt the information that is displayed and how it is further computed. This would clearly be a desirable development in the future.

When considering the rising complexity of data and information in the world, it appears evident that even adapted data visualization cannot be the sole solution to making this information truly accessible for users. Adaptivity may ultimately be understood as providing an individualized information display and decision support. To enable this, a shared semantics between users and systems needs to be developed. Only through teaching the machine how virtual (data) objects relate to each other, the system may be able to provide useful decision support that not only follows a comprehensible logic, but also considers the individual users as cognitive beings that are also prone to typical judgement (and other cognitive) biases. This next step toward individually aiding users in their data-driven decisions can also be considered a step toward artificial intelligence, as we enable machines to apply human (user-centered) perspectives.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The paper was outlined and conceived by TP. All authors contributed to the design of the study. TP organized the dataset, performed the statistical analyses, and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This publication was funded by the German Research Foundation (DFG) and the University of Würzburg in the funding program Open Access Publishing. This work was supported by SAP SE and the University of Würzburg.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00009/full#supplementary-material

Datasheet S1 | Study 1 raw data.

Datasheet S2 | Study 2 raw data.

Datasheet S3 | Study 3 raw data.

Datasheet S4 | Study 4 raw data.

REFERENCES

Ackerman, P., and Beier, M. (2007). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition. *J. Exp. Psychol. Appl*. 13, 249–272. doi: 10.1037/1076-898X.13.4.249

Amar, R., Eagan, J., and Stasko, J. (2005). "Low-level components of analytic activity in information visualization," in *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization* (Minneapolis, MN), 15–21.

Baddeley, A. (1992). Working Memory. *Science* 255, 556–559. doi: 10.1126/science.1736359

Bostock, M., Ogievetsky, C., and Heer, J. (2011). D^3 data-driven documents. *IEEE Trans. Vis. Comput. Graph.* 17, 2301–2309. doi: 10.1109/TVCG.2011.185

Cleveland, W., and McGill, R. (1984). Graphical perception: theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.* 79, 531–554. doi: 10.1080/01621459.1984.10478080

Conati, C., Carenini, G., Toker, D., and Lallé, S. (2015). "Towards user-adaptive information visualization," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (Austin, TX), 4100–4106.

Conati, C., and Maclaren, H. (2008). "Exploring the role of individual differences in information visualization," in *Proceedings of the Working Conference on Advanced Visual Interfaces* (Napoli), 199–206.

Convertino, G., and Echenique, A. (2017). "Self-service data preparation and analysis by business users: new needs, skills, and tools," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, CO), 1075–1083.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., and Wilson, L. (1999). Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia* 37, 1189–1199. doi: 10.1016/S0028-3932(98)00159-6

Demiralp, C., Bernstein, M., and Heer, J. (2014). Learning perceptual kernels for visualization design. *IEEE Trans. Vis. Comput. Graph.* 20, 1933–1942. doi: 10.1109/TVCG.2014.2346978

Eells, W. (1926). The relative mreits of circles and bars for representing component parts. *J. Am. Stat. Assoc.* 21, 119–132. doi: 10.1080/01621459.1926.10502165

Few, S. (2004). *Show Me the Numbers*. Oakland, CA: Analytics Press.

Few, S. (2009). *Now You See It*. Oakland, CA: Analytics Press.

Friel, S., Curcio, F., and Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *J. Res. Math. Educ.* 32, 124–158. doi: 10.2307/749671

Fry, E. (1981). Graphical literacy. *J. Read* 24, 383–389.

Gal, I. (2002). Adult's statistical literacy: meanings, components, responsibilities. *Int. Stat. Rev*. 70, 1–51. doi: 10.1111/j.1751-5823.2002.tb00336.x

Galesic, M., and Garcia-Retamero, R. (2011). Graph literacy: a cross-cultural comparison. *Med. Decis. Making* 31, 444–457. doi: 10.1177/0272989X10373805

Garcia-Retamero, R., Cokely, E., Ghazal, S., and Joeris, A. (2016). Measuring graph literacy without a test: a brief subjective assessment. *Med. Decis. Making* 36, 854–867. doi: 10.1177/0272989X16655334

Germanakos, P. B. (2016). *Human-Centred Web Adaptation and Personalization: From Theory to Practice*. Berlin: Springer Verlag.

Gillian, D., and Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. *Hum. Factors* 36, 419–440. doi: 10.1177/001872089403600303

Gotz, D., and Wen, Z. (2009). "Behavior-driven visualization recommendation," in *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, FL), 315–324.

Grammel, L., Tory, M., and Storey, M.-A. (2010). How information visualization novices construct visualizations. *IEEE Trans. Vis. Comput. Graph.* 16, 943–952. doi: 10.1109/TVCG.2010.164

Harrison, L., Yang, F., Franconeri, S., and Chang, R. (2014). Ranking visualizations of correlation using Weber's law. *IEEE Trans. Vis. Comput. Graph.* 20, 1943–1952. doi: 10.1109/TVCG.2014.2346979

Heer, J., and Bostock, M. (2010). "Crowdsourcing graphical perception: using mechanical turk to assess visualization design," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, GA), 201–212.

Heer, J., Bostock, M., and Ogievetsky, V. (2010). A tour through the visualization zoo. *Commun. ACM* 53, 59–67. doi: 10.1145/1743546.1743567

Heer, J., Kong, N., and Agrawala, M. (2009). "Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations,"in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA), 1303–1312.

Hollands, J., and Spence, I. (1992). Judgments of change and proportion in graphical perception. *Hum. Factors* 34, 313–334. doi: 10.1177/001872089203400306

Howe, B., and Cole, G. (2010). "SQL is dead; long live SQL - lightweight query services for *Ad Hoc* research data," in *Talk Given at 4th Microsoft eScience Workshop* (Berkeley, CA).

Huestegge, L., and Philipp, A. (2011). Effects of spatial compatibility on integration processes in graph comprehension. *Atten. Percept. Psychophys.* 73, 1903–1915. doi: 10.3758/s13414-011-0155-1

Huestegge, L., and Poetzsch, T. H. (2018). Integration processes during frequency graph comprehension: performance and eye movements while processing tree maps vs. pie charts. *Appl. Cog. Psychol.* 32, 200–216. doi: 10.1002/acp.3396

Kaiser, P. (1996). *The Joy of Visual Perception*. Toronto, ON. Available online at: http://www.yorku.ca/eye/ (accessed July 27, 2017).

Kay, M., and Heer, J. (2016). Beyond Weber's law: a second look at ranking visualizations of correlation. *IEEE Trans. Vis. Comput. Graph.* 22, 469–478. doi: 10.1109/TVCG.2015.2467671

Kelava, A., and Schermelleh-Engel, K. (2008). "Latent-state-trait-theorie," in *Testtheorie und Fragebogenkonstruktion,* eds H. Moosbrugger and A. Kelava (Berlin: Springer Verlag), 343–360.

Lallé, S., Toker, D., Conati, C., and Carenini, G. (2015). "Prediction of users' learning curves for adaptation while using an information visualization," in *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, GA), 357–368.

Lee, S., Kim, S.-H., Hung, Y.-H., Lam, H., Kang, Y.-A., and Yi, J. S. (2016). How do people make sense of unfamiliar visualizations? A grounded model of novice's information visualization sensemaking. *IEEE Trans. Vis. Comput. Graph.* 22, 499–508. doi: 10.1109/TVCG.2015.2467195

Legge, G., and Bigelow, C. (2011). Does print size matter for reading? A review of findings from vision science and typography. *J. Vis.* 11, 1–22. doi: 10.1167/11.5.8

Lennerholt, C., van Laere, J., and Söderström, E. (2018). "Implementation challenges of self service business intelligence: a literature review," in *Proceedings of the 51st Hawaii International Conference on System Sciences* (Hawaii, HI), 5055–5063.

Lewandowsky, S., and Spence, I. (1989). Discriminating strata in scatterplots. *J. Am. Stat. Soc.* 84, 682–688. doi: 10.1080/01621459.1989.10478821

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Trans. Graph.* 5, 110–141. doi: 10.1145/22949.22950

Matejka, J., and Fitzmaurice, G. (2017). "Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing," *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Denver, CO), 1290–1294.

Moore, J. (2017). Data visualization in support of executive decision making. *Interdiscipl. J. Inf. Knowledge Manage.* 12, 125–138. doi: 10.28945/3687

North, C., and Shneiderman, B. (2000). Snap-together visualization: can users construct and operate coordinated visualizations? *Int. J. Hum. Comput. Stud.* 53, 715–739. doi: 10.1006/ijhc.2000.0418

Okan, Y., Garcia-Retamero, R., Cokely, E., and Maldonado, A. (2012). Individual differences in graph literacy: overcoming denominator neglect in risk comprehension. *J. Behav. Decis. Mak.* 25, 390–401. doi: 10.1002/bdm.751

Orlov, P., Ermolovy, T., Laptev, V., Mitrofanov, A., and Ivanov, V. (2016). "The eye-tracking study of the line charts in dashboards design," in *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: IVAPP (VISIGRAPP 2016)* (Rome), 205–213.

Oscar, N., Mejía, S., Metoyer, R., and Hooker, K. (2017). "Towards personalized visualization: information granularity, situation, and personality," in *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh), 811–819.

Powers, D. (2012). *Beginning CSS3*. Berkeley, CA: Apress.

Ratwani, R., Trafton, J., and Boehm-Davis, D. (2008). Thinking graphically: connecting vision and cognition during graph comprehension. *J. Exp. Psychol. Appl.* 14, 36–49. doi: 10.1037/1076-898X.14.1.36

Riechelmann, E., and Huestegge, L. (2018). Spatial legend compatibility within versus between graphs in multiple graph comprehension. *Atten. Percept. Psychophys.* 80, 1011–1022. doi: 10.3758/s13414-018-1484-0

Roberts, J. (2007). "State of the art: coordinated & multiple views in exploratory visualization," in *Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization* (Zurich), 61–71.

Roberts, K., Norman, R., Duke, N., Morsink, P., Martin, N., and Knight, J. (2013). Diagrams, timelines, and tables - oh my! fostering graphical literacy. *Read. Teacher* 67, 12–23. doi: 10.1002/TRTR.1174

Sadana, R., and Stasko, J. (2016). "Designing multiple coordinated visualizations for tablets," in *Proceedings of the Eurographics Conference on Visualization* (Groningen).

Scherr, M. (2009). *Multiple and Coordinated Views in Information Visualization*. Unpublished Research Paper. Munich: University of Munich, Germany.

Shah, P., and Hoeffner, J. (2002). Review of graph comprehension research: implications for instruction. *Educ. Psychol. Rev.* 14, 47–69. doi: 10.1023/A:1013180410169

Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.* 11, 92–99. doi: 10.1145/102377.115768

Simkin, D., and Hastie, R. (1987). An information-processing analysis of graph perception. *J. Am. Stat. Assoc.* 82, 454–465. doi: 10.1080/01621459.1987.10478448

Stasko, J., Görg, C., and Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Inf. Vis.* 7, 118–132. doi: 10.1057/palgrave.ivs.9500180

Stolte, C., and Hanrahan, P. (2000). "Polaris: a system for query, analysis and visualization of multi-dimensional relational databases," in *Proceedings of the IEEE Symposium on Information Vizualization, Vol. 8* (Salt Lake City, UT), 5–14.

Strasburger, H., Rentschler, I., and Jüttner, M. (2011). Peripheral vision and pattern recognition: a review. *J. Vis.* 11, 1–82. doi: 10.1167/11.5.13

Toker, D., Conait, C., Steichen, b., and Carenini, G. (2013). "Individual user characteristics and information visualization: connecting the dots through eye tracking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris), 295–304. doi: 10.1145/2470654.2470696

van den Elzen, S., and van Wijk, J. (2013). "Small multiples, large singles: a new approach for data exploration," in *Proceedings of the 15th Eurographics Conference on Visualization* (Leipzig), 191–200.

Vartak, M., Huang, S., Siddiqui, T., Madden, S., and Parameswaran, A. (2017). Towards visualization recommencation systems. *ACM SIGMOD Rec.* 45, 34–39. doi: 10.1145/3092931.3092937

Vartak, M., Rahman, S., Madden, S., Parameswaran, A., and Polyzotis, N. (2015). "SeeDB: efficient data-driven visualization recommendations to support visual analytics," in *Proceedings of the VLDB Endowment* (Kohala Coast, HI), 2182–2193.

Velez, M., Silver, D., and Tremaine, M. (2005). "Understanding visualization through spatial ability differences," in *Proceedings IEEE Visualization* (Minneapolis, MN), 511–518.

Viegas, F., Wattenberg, M., van Ham, F., Kriss, J., and McKeon, M. (2007). ManyEyes: a site for visualization at internet scale. *IEEE Trans. Vis. Comput. Graph.* 13, 1121–1128. doi: 10.1109/TVCG.2007.70577

Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., and Heer, J. (2016). "Towards a general-purpose query language for visualization recommendation," in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics* (San Francisco, CA), 1–6.

Wongsuphasawat, K., Qu, Z., Moritz, D., Chang, R., Ouk, F., Anand, A., et al. (2017). "Voyager 2: augmenting visual analysis with partial view specifications," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, CO), 2648–2659.

Conflict of Interest: TP and PG were employed by the SAP SE.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received no further funding by SAP SE. SAP SE was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/frai.2020.00011

Adapting Learning Activity Selection to Emotional Stability and Competence

Manal Alhathli[1,2], *Judith Masthoff*[1,3]* *and Nigel Beacham*[1]

[1] Department of Computing Science, University of Aberdeen, Aberdeen, United Kingdom, [2] Vice-Rectorate of Development and Quality, Princess Nourah Bint Abdul Rahman University, Riyadh, Saudi Arabia, [3] Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands

This paper investigates how humans adapt next learning activity selection (in particular the knowledge it assumes and the knowledge it teaches) to learner personality and competence to inspire an adaptive learning activity selection algorithm. First, the paper describes the investigation to produce validated materials for the main study, namely the creation and validation of learner competence statements. Next, through an empirical study, we investigate the impact on learning activity selection of learners' emotional stability and competence. Participants considered a fictional learner with a certain competence, emotional stability, recent and prior learning activities engaged in, and selected the next learning activity in terms of the knowledge it used and the knowledge it taught. Three algorithms were created to adapt the selection of learning activities' knowledge complexity to learners' personality and competence. Finally, we evaluated the algorithms through a study with teachers, resulting in an algorithm that selects learning activities with varying assumed and taught knowledge adapted to learner characteristics.

Keywords: learning, adaptation, educational recommender, competency, emotional stability, personalization

OPEN ACCESS

Edited by:
Styliani Kleanthous,
Open University of Cyprus, Cyprus

Reviewed by:
Nathalie Guin,
Université de Lyon, France
Cesar Collazos,
University of Cauca, Colombia

****Correspondence:***
Judith Masthoff
j.f.m.masthoff@uu.nl

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change,
a section of the journal Frontiers in Artificial Intelligence

Received: *30 September 2019*
Accepted: *28 February 2020*
Published: *24 March 2020*

Citation:
Alhathli M, Masthoff J and Beacham N (2020) Adapting Learning Activity Selection to Emotional Stability and Competence. Front. Artif. Intell. 3:11. doi: 10.3389/frai.2020.00011

1. INTRODUCTION

Intelligent Tutoring Systems extend the traditional information-delivery learning system by considering learners' characteristics to improve the effectiveness of a learner's experience (Brusilovsky, 2003). Whilst traditional e-learning has contributed to flexibility in learning and reduced education costs, ITS attempt to fit the particular needs of each individual (Park and Lee, 2004; Brusilovsky and Millán, 2007; Siddappa and Manjunath, 2008; Dascalu et al., 2016).

Adapting ITS to individual learner characteristics helps learners to achieve learning goals and supports personalized learning (Brusilovsky, 1998a,b; Ford and Chen, 2000; Drachsler et al., 2008a; Santos and Boticario, 2010). Several studies have shown that the main problem with traditional e-learning is the lack of learner satisfaction due to delivering the same learning experience to all learners, irrespective of their prior knowledge, experience, and preferences (Ayersman and von Minden, 1995; Cristea, 2003; Rumetshofer and Wöß, 2003; Stewart et al., 2005; Di Iorio et al., 2006; Sawyer et al., 2008). Researchers have tried to address this dissatisfaction by attempting to personalize the learning experience for the learner. A personalized learning experience can help to improve learner satisfaction with the learning experience, learning efficiency, and educational effectiveness (Brusilovsky, 2001; De Bra et al., 2004; Huang et al., 2009). Most research on adaptive learning interaction shows an increase in learning outcomes (Anderson et al., 1995; Vandewaetere et al., 2011).

An important aspect of adaptive e-learning is the adaptive selection of learning activities. In fact, the main goal of adaptive e-learning was identified by Dagger et al. (2005) as "*e-learning content, activities and collaboration, adapted to the specific needs and influenced by specific preferences of the learner and built on sound pedagogic strategies*." Studies have confirmed that the role of adaptation in e-learning is to improve the instruction content given to heterogeneous learner groups (Brusilovsky et al., 1998; Seters et al., 2011). Personalizing the selection of learning activities is needed to make learning more efficient (Camp et al., 2001; Salden et al., 2004; Kalyuga and Sweller, 2005).

Previous studies show that adaptive activity selection impacts factors, such as attitude and behavior (Ones et al., 2007), skills acquisition (Oakes et al., 2001), and productivity (Judge et al., 1999; Bozionelos, 2004). Considering individual differences among learners will improve learning achievement (Shute and Towle, 2003; Tseng et al., 2008). Personalized activity selection yields more effective and efficient learning outcomes, with researchers reporting a positive effect on learners' motivation and learning efficiency (Schnackenberg and Sullivan, 2000; Corbalan et al., 2008).

Learning is influenced by both characteristics of the learner (such as expertise, abilities, attitudes, performance, mental effort, personality) and characteristics of the learning activity (LA) (such as LA complexity, LA type, amount of learner support provided) (Lawless and Brown, 1997; Zimmerman, 2002; Salden et al., 2006; Okpo et al., 2018).

Previously, in six focus group studies (Alhathli et al., 2018b), we investigated what type of LAs to select for a particular learner. Results showed a clear impact of personality (self-esteem, openness to experience, emotional stability) on the use of prior knowledge and topics taught in LA selection. Focus group participants mentioned several other factors that should be considered when selecting a LA, such as a learner's academic record and ability and the LA's difficulty. Given the focus group results, we decided to investigate the impact of Emotional Stability (ES) and learners' competence on the selection of the next LA. In particular, this paper investigates the impact on the selected LA content: both the knowledge taught by the LA and the prior knowledge it uses. The LA style (e.g., visual vs. textual) is not included in this study, as we studied the impact of personality on the selected LA style before (Alhathli et al., 2018a).

2. BACKGROUND AND RELATED WORK

2.1. Learner Characteristics to Adapt to

Researchers have shown an increased interest in adapting to learner characteristics, such as personality traits, motivation, performance, cognitive efficiency, needs, and learning style (Miller, 1991; Wolf, 2003; Shute and Zapata-Rivera, 2007; Schiaffino et al., 2008; Komarraju et al., 2011; Vandewaetere et al., 2011; Richardson et al., 2012; Alhathli et al., 2016, 2017; Dennis et al., 2016; Okpo et al., 2016, 2017). There is considerable debate around which learner characteristics are worth modeling more than others[1]. Vandewaetere et al. (2011) classified individual characteristics into three groups:

- Cognitive, which is a collection of cognition related characteristics, such as the previous knowledge of the learner (Graesser et al., 2007), learners' abilities (Lee and Park, 2008), learning style (Germanakos et al., 2008), and learning objectives (Kelly and Tangney, 2002);
- Affective, which is a collection of feeling related attributes, such as learner mood (Beal and Lee, 2005b), self-efficacy (Mcquiggan et al., 2008), disappointment (Forbes-Riley et al., 2008) and confusion (Graesser et al., 2008); and
- Behavior, whereby a learner behaves differently when they are interacting with computers. These behavioral characteristics can be related to the need for help or feedback (Koutsojannis et al., 2001), the degree of self-regulated learning (Azevedo, 2005), and the number of attempts, tasks and learner experience (Hospers et al., 2003).

In our classification in **Table 1**, we broadened the affective category to psychological aspects, including also personality traits, motivation, and mental effort. We extended the cognitive category to include cognitive style as distinct from learning style. We renamed the behavioral category to include performance. We added an additional category called personal information for learner characteristics not covered by the classification of Vandewaetere et al. (2011), such as demographics and cultural background. **Table 1** shows examples of research into adapting to these learner characteristics.

The learner characteristics investigated in this paper are Personality and Prior knowledge and Competence.

2.1.1. Personality

Previous studies have acknowledged that both personality and general cognitive ability influence learners' performance (Ree et al., 1994; Barrick et al., 2001; Barrick, 2005). It has been suggested that there is a convincing relation between personality and other factors, such as attitude and behavior (Ones et al., 2007), skill acquisition (Oakes et al., 2001), and productivity (Judge et al., 1999; Bozionelos, 2004). Personality can be defined as the individual differences in people's emotional, interpersonal, experiential, attitudinal and motivational styles (John and Srivastava, 1999). Researchers have shown an increased interest in adapting to personality traits (Miller, 1991; Komarraju et al., 2011; Richardson et al., 2012; Dennis et al., 2016; Okpo et al., 2016, 2017).

The most adopted model of personality is the Five-Factor Model (also known as the Big Five), which is based on five dimensions (Costa and McCrae, 1992a, 1995): (i) extroversion, (ii) agreeableness, (iii) conscientiousness, (iv) emotional stability, (v) openness to experience (McCrae, 1992). Extroversion refers to a higher degree of sociability, energy, assertiveness, and talkativeness. Emotional stability refers to the opposite of neurotism, i.e., someone who is calm and not easily upset.

[1]This includes a debate about whether learning styles are a valid construct to consider at all Kirschner (2017). Our own research in Alhathli et al. (2017) in fact showed little impact of learning styles.

TABLE 1 | Examples of learner characteristics adapted to in adaptive educational systems.

Learner characteristics adapted to		Example research
	Personality; Big 5	Robison et al., 2010; Nunes and Hu, 2012; Alhathli et al., 2016, 2018b; Dennis et al., 2016
Psychological aspects	Self-efficacy	Beal and Lee, 2005a; Mcquiggan et al., 2008
	Mental effort	Salden et al., 2006; Okpo et al., 2018
	Motivation	Beal and Lee, 2005a; Cocea and Weibelzahl, 2006
	Affective state and mood	Beal and Lee, 2005b; Forbes-Riley et al., 2008; Graesser et al., 2008; Odo et al., 2018
Behavior and performance	Learner competence	Mitrović et al., 1996; Davidovic et al., 2003; Tsiriga and Virvou, 2004; Cheng et al., 2008; Corbalan et al., 2008
	Problem solving skills	Melis et al., 2001; Pholo and Ngwira, 2013
	Help seeking behavior and self-regulated learning	Koutsojannis et al., 2001; Azevedo, 2005
	Learner progress	Brusilovsky et al., 1996; Revilla et al., 2008; Trotman and Handley, 2008; Verdú et al., 2012
Cognition	Cognitive style	Triantafillou et al., 2004; Graesser et al., 2007; Mampadi et al., 2011; Lo et al., 2012; Alhathli et al., 2018a
	Knowledge state and prior knowledge	Shute, 1995; Ray and Belden, 2007; Kelly, 2008; Petrovica, 2013
	Learning style	Magoulas et al., 2003; Sun and Cheng, 2007; Germanakos et al., 2008; Latham et al., 2012; El-Bishouty et al., 2014; Alhathli et al., 2017
	Learner objectives	Kelly and Tangney, 2002; Vassileva and Bontchev, 2006
Personal information	Learner profile, demographics, cultural background, and preferences	Hwang, 1998; Widyantoro et al., 1999; Chang et al., 2000; Sugiyama et al., 2004; Reategui et al., 2008; Adamu Sidi-Ali et al., 2019

Openness to experience refers to those who are interdependent-minded, and intellectually strong. Conscientiousness refers to being disciplined, organized, and achievement-oriented. Finally, Agreeableness refers to being good-natured, helpful, trustful, and cooperative (Miller, 1991). These traits have been found across all cultures (McCrae and Costa, 1997; Salgado, 1997). In addition, these traits are relatively stable over time (Costa and McCrae, 1992b).

Several studies have shown the effect of personality on the learning process, and it has been investigated that certain personality traits consistently correlate with learner achievement, motivation, and success (Komarraju and Karau, 2005; Poropat, 2009; Clark and Schroth, 2010; Komarraju et al., 2011; Hazrati-Viari et al., 2012; Richardson et al., 2012).

2.1.2. Prior Knowledge and Competence

Numerous terms have been used to refer to prior knowledge (e.g., current knowledge, expert knowledge, personal knowledge, and experiential knowledge) (Dochy, 1992, 1994). Interest in a learner's prior knowledge has appeared in many educational studies. An individual's prior knowledge is considered as a set of skills, or abilities that are present in the learning process (Jonassen and Grabowski, 1993; Shane, 2000). Previous investigations have demonstrated the potential impact of prior knowledge on cognitive processes, with positive and significant effects on learner's performance, abilities, and achievement (Byrnes and Guthrie, 1992; Dochy, 1994; Gaultney, 1995; Thompson and Zamboanga, 2003).

In our focus groups, we found that prior knowledge impacts the selection of the next LA. Thus, we decided to use learners' competence in terms of learner knowledge and ability.

Competence can be defined differently depending on the discipline. The dictionary defines competence as a condition or as quality of effectiveness. Competence refers to an individual's capability, sufficiency, ability, and successes. A large amount of competence research refers to the skills and requirements needed for a particular task or profession (Willis and Dubin, 1990; Parry, 1996). Competence is seen as a reflection of multiple concepts, such as performance. Competence and performance are related, with competence depicting the mean of better performance (Klemp, 1979; Woodruffe, 1993).

However, performance can be affected by other factors, such as motivation and effort (Schambach, 1994). Competencies are also considered as a core component of goal achievement. Achievement goals are defined as a cognitive representation of a competence efficiency and ability that an individual seeks to obtain (Elliot, 1999; Bong, 2001; Elliot and McGregor, 2001). Competence can be defined depending on the standard or referent that is used in evaluation (Elliot and Thrash, 2001). Competence may be evaluated according to three different standards, as follows: (1) absolute, the requirement of the task itself; (2) intra-personal, past or maximum attainment; and (3) normative, the performance of others (Butler, 1988; Elliot and McGregor, 2001).

2.2. Educational System Characteristics to Adapt

Many aspects of an educational system can be adapted to a learner. For example, Masthoff (1997) argued for adapting navigation through the course content, exercise selection, feedback, instructions, provision of hints, and content presentation. For example, feedback has been adapted to performance and personality (Dennis et al., 2016) and culture (Adamu Sidi-Ali et al., 2019), difficulty level to performance, personality and effort (Okpo et al., 2018), navigational control to learner goals and knowledge (Masthoff, 1997), and learning content and presentation to learning styles (Bunderson and Martinez, 2000).

This paper focuses on adaptive learning activity selection. **Table 2** provides examples of adaptive educational systems that include adaptive LA selection, the learner characteristics used to guide the adaptation, and the system control used to provide the adaptation. The following types of system control can be distinguished:

(1) *Curriculum Sequencing* provides learners with a planned sequence of learning contents and tasks (Brusilovsky, 2003),
(2) *Adaptive Navigation Support* helps learners to find their paths in the learning contents according to the goals, knowledge, and other characteristics of an individual learner (Brusilovsky, 1996),
(3) *Collaborative Filtering* (and other educational recommender systems' techniques) supports learners to find learning resources that are relevant to their needs and interests (Recker and Walker, 2003; Recker et al., 2003; Schafer et al., 2007; Drachsler et al., 2008a),
(4) *Adaptive Presentation* supports learners by providing individualized content depending on their preferences, learning style and other information stored in the learner model (Beaumont and Brusilovsky, 1995).

The LA selection in this paper is concerned with selecting activities that are well-suited to learners' personality, prior knowledge and competence. This is related to Adaptive Navigation and Educational Recommender Systems, given a LA is selected as most suitable for a learner based on the knowledge the LA assumes and teaches. The LA selected by the system can be used to support learners in finding what LA to do next or by an ITS to automatically present that LA.

The domain model in our studies contains the LAs, and in particular their topics and the type and quantity of knowledge they use and produce.

2.2.1. Learning Activity Topic

Educational systems which recommend or provide personalized learning contents often require information about the topics covered in the learning materials, courses, and assignments it selects from (Liang et al., 2006; Soonthornphisaj et al., 2006; Prins et al., 2007; Yang and Wu, 2009; Ricci et al., 2011). Bloom's Taxonomy defines three overarching domains of LAs: Cognition (e.g., teaching mental skills), Affective (e.g., teaching attitudes), and Psychomotor (teaching manual of physical skills) (Bloom, 1956). This paper focuses on the cognitive domain. Within the cognitive domain, there are many sub-domains. For example, educational recommender systems have been developed for programming (Mitrovic et al., 2002; Wünsche et al., 2018) and learning languages (Hsu, 2008; Wang and Yang, 2012). Even within such a sub-domain, multiple topics exist. For example, when teaching somebody English, one could have a LA on ordering food, and a different activity on buying groceries. Educational recommender systems often select based on learner interests, so need detailed information on the topics covered in a LA.

TABLE 2 | Examples of adaptive educational systems that adapt content selection.

System	Learner characteristics	System control
CDG (Vassileva, 1997)	Personal traits; learning goal; preferences	Curriculum sequencing
AST (Specht et al., 1997)	Knowledge level; learning style preferences	Curriculum sequencing
KBS hyperbook (Henze et al., 1999)	Knowledge level; learning goals	Adaptive navigation support
Arthur (Gilbert and Han, 1999)	Learning style preferences	Curriculum sequencing
Altered Vista (Recker and Walker, 2003)	Preferences	Collaborative filtering
RACOFI (Anderson et al., 2003)	Multidimensional ratings	Collaborative filtering
INSPIRE (Papanikolaou et al., 2001)	Knowledge level; learning style	Adaptive presentation
Learning object sequencing (Shen and Shen, 2004)	Knowledge base; learner competence	Curriculum sequencing
QSIA (Rafaeli et al., 2004)	Knowledge sharing	Collaborative filtering
Rmashed (Drachsler et al., 2008b)	Learning goals	Collaborative filtering
CYCLADES (Avancini and Straccia, 2005)	User interests; preferences	Collaborative Filtering
Rmashed (Drachsler et al., 2008b)	Learning goals	Collaborative filtering
iLearning (Wang et al., 2014)	Knowledge level	Collaborative filtering

2.2.2. Learning Activity Knowledge

Incorporating learner characteristics, such as the learner's knowledge, interests and goals in an adaptive educational system is a well-established approach discussed by Brusilovsky (2003, 2007). To adapt LA selection, a match needs to be made between the learner's knowledge, goals and interests and what LAs have to offer and require. In traditional education, LAs are often described in terms of prerequisites (the knowledge required of a learner to participate in a LA) and learning outcomes (the knowledge the learner will gain by successfully completing a LA) (Anderson et al., 2001).

3. CREATION AND VALIDATION OF LEARNER COMPETENCE STATEMENTS

This section describes the development and validation of competence statements used in later studies. Many statements

TABLE 3 | Competence statements (grouped by initial categorization) mapped to competence rating.

Initial		Competence rating by participants %										Average	
Cat.	**Statement**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Rating**	**Median**
A	No	95%	5%									1.06	1.00
	Very low	28%	72%									1.71	2.00
	Poor	17%	50%	28%	5%							2.22	2.00
	Hardly any	28%	50%	11%			11%					2.28	2.00
B	Little		22%	50%	28%							3.06	3.00
	Low	5%	11%	62%	17%	5%						3.06	3.00
	Limited	5%	17%	28%	34%		11%			5%		3.72	3.50
	Slight		5%	11%	39%	28%	17%					4.39	4.00
	Some			23%	39%	5%	28%		5%			4.61	4.00
C	Fair	5%		5%	5%	28%	39%	18%				5.33	6.00
	Quite some				28%	17%	22%	22%	11%			5.72	6.00
	Medium					56%	22%	17%	5%			5.72	5.00
	Moderate				11%	39%	17%	28%		5%		5.83	5.50
	Standard					39%	17%	22%	22%			6.28	6.00
D	Good				5%	28%		34%	33%			6.61	7.00
	Sufficient					5%	28%	56%	11%			6.72	7.00
	Recognized				5%	17%	17%	34%	5%	5%	17%	6.76	7.00
	Much				5%	5%	5%	18%	28%	39%		7.00	8.00
	Very good				5%		16%	11%	28%	40%		7.72	8.00
	High							28%	50%	17%	5%	8.00	8.00
	Advanced						5%	23%	33%	28%	11%	8.17	8.00
E	Very high						5%	17%	5%	34%	39%	8.83	9.00
	Excellent								22%	22%	56%	9.23	10.00
	Full								16%	28%	56%	9.39	10.00
	Outstanding								5%	39%	56%	9.50	10.00
	Extreme						5%				95%	9.78	10.00

Statements in bold were used in the follow-on study.

can be used to describe different levels of competency, but no existing list clearly defined varying levels of individual competence. Initially, 26 statements were produced to cover five categories of learners' competence. All statements are commonly used to depict different competence levels. **Table 3** shows the resulting statements and their initial categorization. These statements will be used in our investigations on the impact of personality and competence on the selection of LA.

3.1. Study Design

3.1.1. Participants

Thirty participants (staff and students of the university) completed an on-line survey (7% aged 18–25, 53% 26–35, 40% 36–45), which took about 15 min to complete. The data from 18 participants were used for the final analysis (9 male, 9 female). The others were excluded due to the low quality of their responses: either straight-lining (giving the same answer to all statements), or not putting "No competence" toward the bottom of the scale as directed in the explanation.

3.1.2. Statement Validation

Participants were shown 26 statements, and rated how much they felt these statements reflect the individual competence of a learner from 1 (no competence at all) to 10 (maximum competence). The order of the competence statements was randomized for each participant. **Table 3** shows the percentage of participants who mapped a statement to a particular number.

3.2. Results

Table 3 shows the percentage of participants who mapped a statement to a particular number. Some statements (e.g., "limited," "slight") showed little agreement between participants, whilst others showed better agreement. We decided to use three statements (shown in bold) for the main studies, which are *"very low," "moderate,"* and *"outstanding."* These statements were selected to ensure a spread of learners' competence, good agreement between participants, and based on the average ratings and the median. However, we decided to exclude "no competence" as it was used in the explanation of the scale that participants saw in the validation experiment, and we excluded "Medium competence" and "Extreme competence" as they could be affected by a comparison with other learners in the class. More statements could be used in future studies; for example "Little competence" (or "Low competence") and "High competence" could be used if one needed five competence statements. For our future studies, we needed only three.

Id	Uses knowledge				Teaches
	Old		Recent		
	A	B	D	E	
1	×				F
2	×				F and B
3	×				F and D
4	×				F and G
5	×	×			F
6	×	×			F and D
7	×	×			F and G
8	×		×		F
9	×		×		F and B
10	×		×		F and E
11	×		×		F and G
12			×		F
13			×		F and B
14			×		F and E
15			×		F and G
16			×	×	F
17			×	×	F and B
18			×	×	F and G

FIGURE 1 | LAs as shown to participants.

4. ADAPTING THE SELECTION OF LEARNING ACTIVITY KNOWLEDGE COMPLEXITY TO EMOTIONAL STABILITY AND COMPETENCE

In this study, we investigate the impact of learners' emotional stability and competence on the selection of the next LA. In particular, we investigate the impact on the selection of both the knowledge taught by the LA and the prior knowledge it uses. We use three levels of competence: "very low," "moderate," and "outstanding." Through an empirical study, we investigate how humans select the next LA for a learner with various levels of emotional stability and competence. Participants considered a fictional learner with certain levels of competence and emotional stability, recent and prior LA engaged in, and selected the next LA in terms of the knowledge it used and the knowledge it taught.

4.1. Variables

The independent variables used for the study are: *Personality Trait Story:* Participants were shown a story about a learner which portrayed a personality trait. Two stories were used depicting Emotional Stability (ES) at either a low or high level. The ES stories were developed and validated by Dennis et al. (2012), Smith et al. (2019). *Learner competence:* Three levels of competence were used: very low, moderate, and outstanding. The dependent variable for the studies is *Learning activity selected:* Participants were shown a table with each row containing a LA (numbered from 1 to 18). For each LA, the table showed the PRE knowledge the LA uses, with a distinction made between old knowledge (topics A and B) and recent knowledge (D,E). It also showed the POST knowledge the LA teaches; this could contain new (F and G), old (A,B), or recent knowledge (D,E). For example, for LA 9, it indicated that it uses old knowledge A and recent knowledge D, and teaches F and B. The LAs available for selection are showed in **Figure 1**. Participants selected one LA, and in doing so made a choice for PRE and POST knowledge. We will use PRE and POST as the dependent variables.

4.2. Procedure

Participants had to pass an English fluency test (Cloze test, Taylor, 1953), to ensure that they could understand the study. Then, they were shown a description of the LA table with examples and two verification questions to ensure they had understood what the table indicated. Next, they were shown six scenarios with different learner competence *(very low, moderate, and outstanding)*, three scenarios depicting *Josh* who was high on ES and another three depicting *James* who was low on ES. They were told that the learner has previously learned topics A and B, and recently finished a LA which taught topics D and E. For each scenario, they selected the next LA for that learner from the table described before *e.g., "Which learning activity would you give to Josh to do next, if you know his competence in both old and recent knowledge is 'Very low'?"* Next, they rated how much they think the selected LA is suited to that learner on a scale from 1 (Not at all) to 5 (Totally suited), and to what extent the selected LA would be enjoyable, would increase skills and confidence (on a scale from 1 strongly disagree to 5 strongly agree).

4.3. Participants

Fifty-three participants responded to the on-line survey. 24 responses were excluded from the study either because they did not pass the English test, or answered the verification questions incorrectly. Twenty-nine participants successfully completed the study (16 female, 13 male; 2 aged 18–25, 11 aged 26–40, 10 aged 36–45, 6 over 46; 8 were students, 19 were teachers, 2 were trainee-teachers).

4.4. Hypotheses

We hypothesized that:

- **H1**. Participants will select different LAs for high ES than low ES learners:
  - **H1.1**. They will select LAs with less complicated POST for low ES than high ES.
  - **H1.2**. They will select LAs with less complicated PRE for low ES than high ES.
- **H2**. Participants will select different LAs for learners with different competence levels:

- **H2.1**. They will select LAs with less complicated POST for lower levels of competence.
- **H2.2.** They will select LAs with less complicated PRE for lower levels of confidence.

- **H3**. There will be an interaction between ES and competence on LA selection.
- **H4**. Participants will rate the suitability of a selected LA and the extent to which it increases confidence differently depending on the PRE and POST. In particularly, we expect that for low ES:
  - **H4.1**. They will rate the suitability of a selected LA higher when it has less complicated POST.
  - **H4.2**. They will rate the suitability of a selected LA higher when it has less complicated PRE.
  - **H4.3**. They will rate the extent to which the selected LA increases confidence higher when it has less complicated POST.
  - **H4.4**. They will rate the extent to which the selected LA increases confidence higher when it has less complicated PRE.

4.5. Results

4.5.1. Initial Observations on LA Selection

Figure 2 shows the proportion of participants who selected a particular LA. The LAs available for selection are summarized in the second and third row of the figure, where PRE indicates the topics the LA uses, and POST indicates the topics the LA teaches. To make the results easier to read, we use more meaningful codes here instead of the A-G participants saw. For PRE we use: (O) only one old topic, (2O) two old topics, (OR) one old and one recent topic, (R) only one recent topic, and (2R) two recent topics. For POST we use: (N) one new topic, (NO) one new topic and one old topic, (NR) one new and one recent topic and (2N) two new topics. For example, the figure shows that 7% selected a LA with PRE knowledge O and POST knowledge N for the very low competence and high ES learner. From the figure, we observe the following:

- Very low competence: Participants tended to select LAs that required old knowledge (O, 2O, OR) for both ES levels. However, participants tended to select LAs that involved learning a combination of new and old knowledge (NO) for high ES, and more just new knowledge (N) for low ES.
- Moderate competence: Participants tended to select LAs that required old knowledge for both levels of ES, but mainly OR, with O and 2O not selected much. Interestingly, a higher proportion of participants selected LAs teaching NR or 2N for high ES, whilst more selected N for low ES.
- Outstanding competence: Participants selected LAs that involved less knowledge to learn (N vs. 2N) for low ES compared to high ES.

So, overall, there is evidence of participants changing their LA selection based on ES, and were indeed selecting LAs with less complicated POST for the low ES learner. This supports hypothesis H1.1. We do not find support here for H1.2.

There is also evidence in support of hypotheses H2.1 and H2.2, as **Figure 2** clearly shows that the proportion of participants selecting more complicated PRE (particularly 2R) and more complicated POST (particularly 2N) increased with an increase in competence.

4.5.2. Impact of ES and Competence on PRE, POST

For the statistical analysis, we coded PRE and POST in such a way that a higher number indicates more (complicated) knowledge. For PRE, we coded O=1, 2O=2, R=2, OR=3 and 2R=4, so assigning higher numbers the more (complicated) knowledge is used[2]. For POST, we coded N=1, NO=2, NR=3, 2N=4, so assigning higher numbers the more (complicated) knowledge was taught. **Figures 3–5** show the overall impact of ES and competence on PRE and POST.

1. Emotional Stability: There was a significant main effect of ES on POST [$F_{(1,168)} = 12.3$, $p < 0.005$][3], but not on PRE. LAs selected for high ES taught more new knowledge than LAs selected for low ES. **Figure 3** shows a trend for LAs with more PRE being selected for high ES than low ES. However, the difference was small and not significant. This support hypothesis H1.2 but not H2.2.
2. Competence: There was a significant main effect of competence on both PRE and POST [$F_{(2,168)} = 6.0$, $p < 0.005$; $F_{(2,168)} = 22.7$, $p < 0.0005$, respectively][4]. For POST, pairwise comparisons showed a significant difference between "very low" and "moderate" competence on the one hand, and "outstanding" competence on the other ($p < 0.0005$), with LAs with more POST selected for "outstanding" competence (mean difference = 1.24 and 1.09, respectively). For PRE, there was a significant difference only between "very low" and "outstanding" competence ($p < 0.005$), with more PRE selected for "outstanding" competence (mean difference = 0.53) (see **Figure 4**). This supports hypotheses H2.1 and H2.2.
3. Interaction between ES and competence: **Figure 5** shows the PRE and POST per competency level for high and low ES. There was no significant interaction effect between ES and competence, so there is no evidence in support of H3.

4.5.3. Suitability, Enjoyment, Increasing Skills, and Confidence

Figure 6 shows participants' suitability ratings for the most selected LAs. **Table 4** shows participants' enjoyment, skills and confidence ratings for the most selected LAs for the different levels of competence and ES. Overall, there were no significant effects of ES and competence on suitability. There was a significant effect of ES on enjoyment [$F_{(1,168)} = 9.6$, $p < 0.005$] with a higher enjoyment rating for high ES (mean of 3.7 compared to 3.3), but not on skills and confidence. There was also a significant effect of competence on enjoyment [$F_{(2,168)} = 4.3$, $p < 0.05$], with a higher enjoyment rating for higher competence (mean of 3.8 for outstanding competence compared

[2]As it is hard to say whether 2O or R requires more knowledge, we coded them the same.
[3]A similar significant effect was found using a non-parametric test.
[4]Similar significant effects were found using non-parametric tests.

Very low competence																		
PRE	O				2O			OR				R				2R		
POST	N	NO	NR	2N	N	NR	2N	N	NO	NR	2N	N	NO	NR	2N	N	NO	2N
High ES	7%	14%			3%	4%		7%	21%	3%	17%			3%		7%	14%	
Low ES	10%	4%			7%	14%		24%	14%	4%	3%	7%		3%		3%	7%	
Moderate competence																		
PRE	O				2O			OR				R				2R		
POST	N	NO	NR	2N	N	NR	2N	N	NO	NR	2N	N	NO	NR	2N	N	NO	2N
High ES				3%	3%	7%	4%	21%	7%	14%	10%	3%	3%	4%	4%	7%	7%	3%
Low ES			3%			4%	3%	24%	17%	4%	7%	14%	3%	7%		10%	4%	
Outstanding competence																		
PRE	O				2O			OR				R				2R		
POST	N	NO	NR	2N	N	NR	2N	N	NO	NR	2N	N	NO	NR	2N	N	NO	2N
High ES				3%			7%	3%		4%	31%	4%			10%		7%	31%
Low ES					3%		4%	3%	3%	7%	28%		4%		10%	21%	7%	10%

FIGURE 2 | Participants' LA selection.

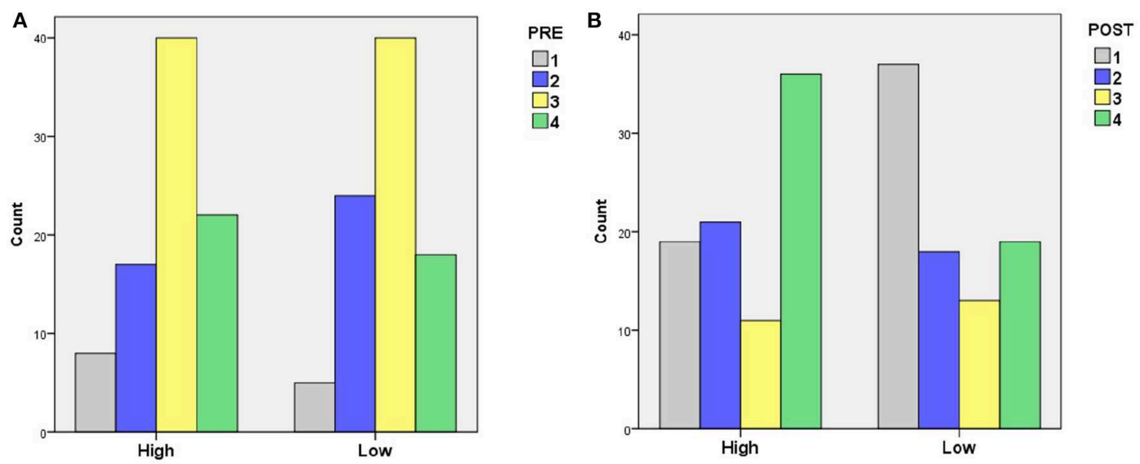


FIGURE 3 | The impact of ES on **(A)** PRE and **(B)** POST knowledge.

to 3.5 for moderate and 3.3 for very low), but not on skills and confidence. There were no significant interaction effects. Given participants' selection of LAs (and hence the LAs for which suitability, enjoyment, skills and confidence were rated) differed based on competence and ES, we also explored this in more detail, though the number of participants is too low for statistical tests.

4.5.3.1. Very low competence

For high ES, the LA which uses more recent knowledge (2R) was rated more suitable than those that used more old knowledge (O and OR) with the same POST (NO). The skills rating for this LA was also higher, whilst it confidence rating was lower. Participants may have felt that the high ES learner did not require a LA that would increase their confidence but rather their skills. For low ES, the LA that teaches less knowledge (N instead of NO) with the same PRE (OR) was rated more suitable. This LA also had a higher rating for confidence, which may mean that participants felt the low ES learner needed to gain more confidence.

4.5.3.2. Moderate competence

For low ES, the LA which uses less knowledge (R) was rated more suitable than the one that uses more knowledge (OR) with the same POST (N). This LA also had a much higher rating for confidence. For high ES, the LA that teaches NR was rated more suitable than the ones teaching N or 2N, with the same PRE (OR). This LA also rated higher on the other aspects.

4.5.3.3. Outstanding competence

For low ES, the LA which teaches less knowledge (N) was rated more suitable than the one that teaches more knowledge (2N)

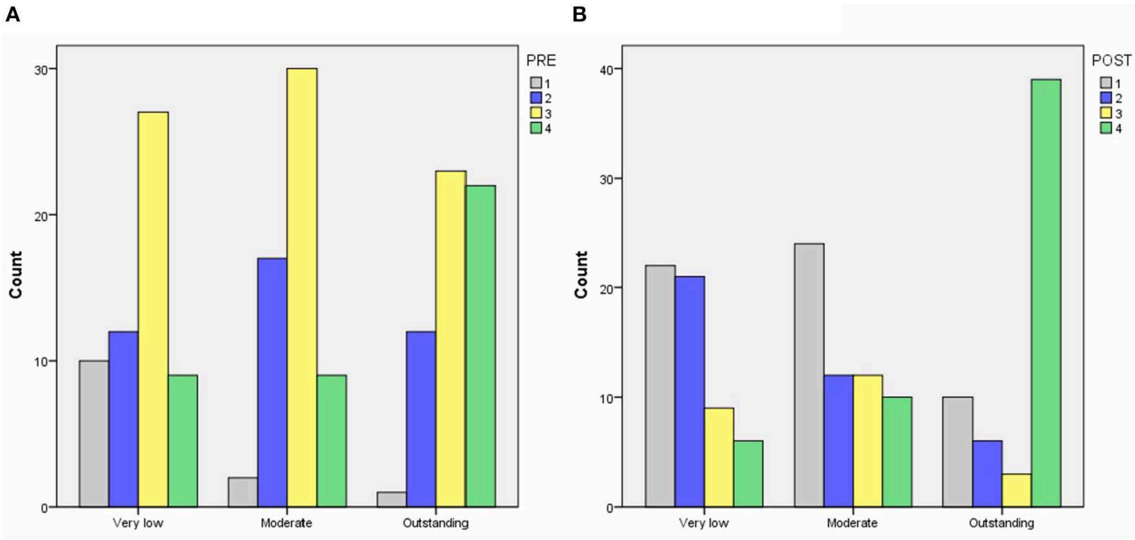


FIGURE 4 | The impact of competence on **(A)** PRE and **(B)** POST knowledge.

with the same PRE (2R). This LA also had a higher rating for confidence, whilst it had a lower rating for skills.

Overall, this seems to suggest that LAs that are teaching less knowledge or using less knowledge are seen as more suitable for low ES, because they may increase confidence. This provides some support for hypotheses H4.1–H4.4.

4.5.4. Initial Algorithms for Adapting Learning Activity Selection Based on the Data

The main concern of this paper was to investigate how to select the next LA for a learner with a particular level of ES and competence. Using the data presented in **Figure 2**, three initial approaches were used to produce algorithms for selecting LAs:

1. *Most frequently chosen LA.* For each combination of competence and ES, we considered which LA was most frequently selected (see summary in **Table 5**). In case of outstanding competence and high ability, two LAs were chosen as often. In this case we selected the one with the same PRE as had been selected for low ES, given there had not been a significant effect of ES on PRE. This resulted in Algorithm 1.
2. *LA produced by combining the most frequently chosen PRE and the most frequently chosen POST.* For each combination of competence and ES, we considered which PRE and which POST were most frequently selected (see summary in **Table 5**). Using the LA which combines the most frequently selected PRE and the most frequently selected POST produced the same results as using the most frequently selected LA[5]. Hence, Algorithm 1 is already in line with the outcome of this approach and no new algorithm was produced.
3. *Top 3 LA exhibiting the largest increase in selection compared to the opposite ES case.* The differences in frequency between the most selected LAs and the second (or even third) most selected LAs tended to be relatively small. Therefore, we also considered for each combination of competence and ES, which top 3 LA showed the largest increase in frequency of selection compared to the opposite ES case. For example, for outstanding competence and high ES, 2R→2N is the top 3 LA which the largest increase in frequency (31% for high ES and only 10% for low ES). This resulted in Algorithm 2.

In the next section, a more complicated statistical approach will be used resulting a third algorithm, and the three algorithms will be evaluated below.

4.5.5. Regression Analysis and Resulting Algorithm

Using the data of the study, two cumulative odds ordinal logistic regressions with proportional odds were run to predict the PRE and POST based on ES and competence[6]. The final model for both PRE and POST statistically significantly predicted the PRE and POST level over and above the intercept-only models [$\chi^2_{(2)} = 10.458$, $p < 0.01$; $\chi^2_{(2)} = 41.759$, $p < 0.0005$, respectively]. The odds ratio of selecting a higher POST level for learners with high ES vs. low ES is 2.707 (95% CI, 1.531–4.787), a statistically significant effect, Wald $\chi^2_{(1)} = 11.740$, $p < 0.005$. This supports hypothesis H1.1. An increase in competence was associated with selecting a higher POST level with an odds ratio of 2.768 (95% CI, 1.919–3.983), Wald $\chi^2_{(1)} = 29.734$, $p < 0.0005$. This supports H2.1 and also provides evidence that competence has slightly more impact on POST than ES. An increase in competence was also associated with selecting a higher PRE level with an odds ratio of 1.756 (95% CI, 1.240–2.487), Wald $\chi^2_{(1)} = 10.059$, $p < 0.005$. This supports H2.2. The odds-ratio for high ES vs. low ES for PRE was not significant, so there is again no evidence of H1.2.

[5]Similarly to the discussion above, for outstanding competence and high ES, the most frequently selected PRE could also have been 2R instead of OR.

[6]ES was used as a factor. Competence was used as a ordinal co-variate, with competence coded 1-3 for very low till outstanding.

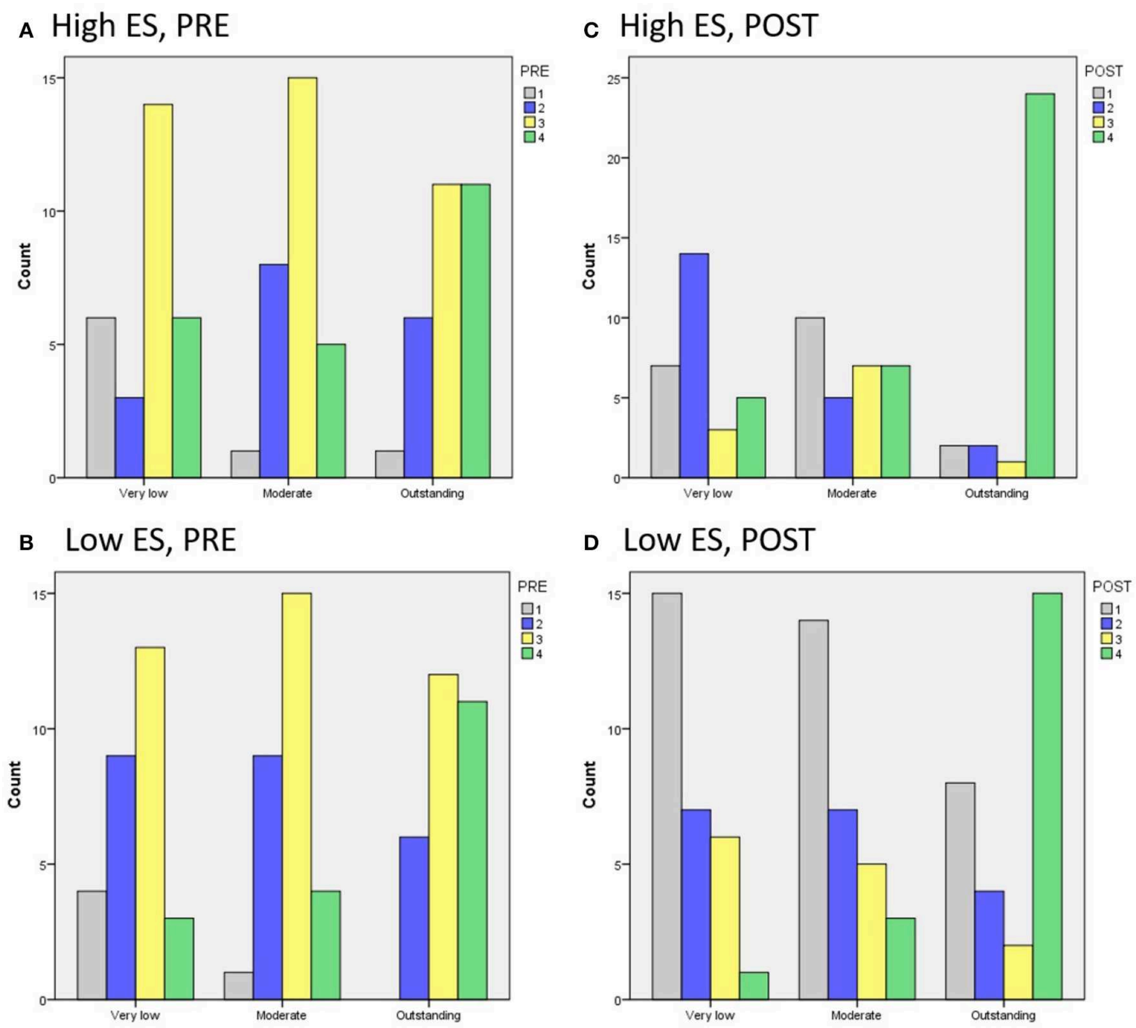


FIGURE 5 | The impact of competence on PRE for **(A)** high and **(B)** low ES, and on POST for **(C)** high and **(D)** low ES.

The model is using an interaction between ES and competence, so provides some support for H3.

The model provides coefficients to calculate a value, as well as thresholds to compare the calculated value against to produce cumulative odds for PRE and POST levels.

The model's coefficients result in the following formulae to calculate Value for PRE and POST:

- PRE:
  - 0.563 X Competence + 0.175 if ES = High
  - 0.563 X Competence if ES = Low
- POST:
  - 1.018 X Competence + 0.996 if ES = High
  - (1.018 X Competence) if ES = Low

The thresholds lead to the following formulae to calculate the natural logarithm of the cumulative odds for PRE and POST:

- PRE:
  - $\ln(\text{Odds}(\text{PRE} \leq 1)) = -1.378$ –Value
  - $\ln(\text{Odds}(\text{PRE} \leq 2)) = 0.381$ –Value
  - $\ln(\text{Odds}(\text{PRE} \leq 3)) = 2.471$ –Value
- POST:
  - $\ln(\text{Odds}(\text{POST} \leq 1)) = 1.572$ –Value
  - $\ln(\text{Odds}(\text{POST} \leq 2)) = 2.643$ –Value
  - $\ln(\text{Odds}(\text{POST} \leq 3)) = 3.386$ –Value

Using these formulae, for each combination of competence and ES we calculated:

- Value, see **Table 6**
- Odds(PRE $\leq$ d), for all PRE levels d
- Probability P(PRE $\leq$ d) for all PRE levels d
- P(PRE=d) for all PRE levels d, using that P(PRE $\leq$ 1) = P
- (PRE = 1) and P(PRE = d+1) = P(PRE $\leq$ d+1) –P(PRE $\leq$ d)

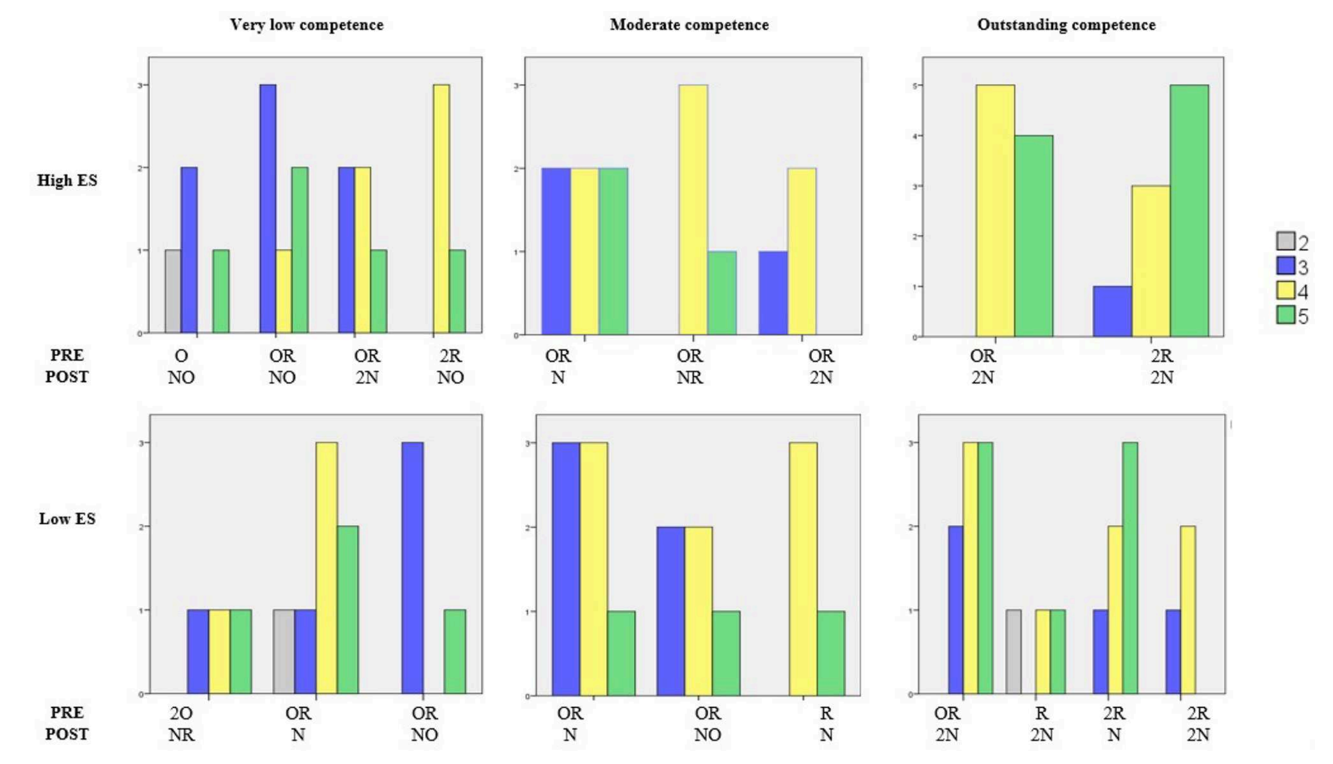


FIGURE 6 | Suitability ratings of the most selected LAs.

TABLE 4 | Mean (stdev) appreciation.

Competence	ES	Percentage (%)	PRE	POST	Enjoyable	Skills	Confidence
Very low	High	21	OR	NO	3.67 (0.81)	3.83 (0.40)	3.83 (0.75)
		17	OR	2N	3.40 (1.51)	3.60 (0.89)	3.60 (0.89)
		14	O	NO	3.00 (81)	3.75 (0.50)	3.50 (0.57)
		14	2R	NO	3.50 (1.00)	4.00 (0.00)	3.25 (0.95)
	Low	24	OR	N	3.14 (1.34)	4.29 (0.48)	3.71 (1.11)
		14	2O	NR	2.25 (1.25)	3.50 (0.57)	3.75 (0.50)
		14	OR	NO	3.00 (0.81)	3.50 (0.57)	3.00 (0.81)
Moderate	High	21	OR	N	3.50 (0.83)	4.00 (0.63)	3.17 (0.75)
		14	OR	NR	4.25 (0.50)	4.25 (0.50)	4.00 (0.81)
		10	OR	2N	3.33 (0.57)	4.00 (0.00)	3.67 (0.57)
	Low	24	OR	N	3.14 (0.69)	4.14 (0.69)	3.00 (0.81)
		14	OR	NO	3.20 (1.09)	3.60 (0.89)	3.60 (0.89)
		14	R	N	4.25 (0.95)	4.50 (0.57)	5.00 (0.00)
Outstanding	High	31	OR	2N	4.11 (0.60)	4.78 (0.44)	4.33 (0.70)
		31	2R	2N	4.22 (0.83)	4.44 (0.72)	4.00 (0.86)
	Low	28	OR	2N	3.14 (0.69)	4.14 (0.69)	3.00 (0.81)
		21	2R	N	3.20 (1.09)	3.60 (0.89)	3.60 (0.89)

- Median PRE m such that $P(PRE \leq m) \geq 0.5 \wedge P(PRE \geq m) \geq 0.5$.

Similar calculations were performed for POST. **Table 6** shows the calculated values for all our combinations of competence and ES for PRE and POST, respectively, and how these values map onto the median PRE and POST levels. The predicted median PRE and POST levels were used to produce Algorithm 3.

This study investigated the impact of learner personality (emotional stability) and competence on the selection of a LA based on the knowledge it uses and the knowledge it teaches. ES and competence both impacted the selection of LAs. There were

TABLE 5 | Most frequently selected LA, PRE, and POST, and the percentage of participants who selected them, and LA with largest increase in selection.

Competence	ES	LA (%)	PRE (%)	POST (%)	Largest increase LA
Very low	High	OR→NO (21%)	OR (48%)	NO (49%)	OR→2N
	Low	OR→N (24%)	OR (45%)	N (37%)	OR→N
Moderate	High	OR→N (21%)	OR (52%)	N (34%)	OR→NR
	Low	OR→N (24%)	OR (52%)	N (48%)	R→N
Outstanding	High	OR→2N (31%), 2R→2N (31%)	OR (38%), 2R (38%)	2N (75%)	2R→2N
	Low	OR→2N (28%)	OR (41%)	2N (52%)	2R→N

Algorithm 1: LA selection based on the most frequent LA selected

```
Input: Emotional stability the learner's level of emotional
       stability; competence the learner's competence level
Output: PRE; POST
begin
    PRE := OR;
    switch Competence do
        case very low do
            if Emotional stability = low then
                POST := N;
            else
                POST :=NO;
            end
        end
        case moderate do
            POST := N;
        end
        case outstandin g do
            POST := 2N;
        end
    end
end
```

Algorithm 2: LA selection based on the largest increase in frequency

```
Input: Emotional stability the learner's level of emotional
       stability; competence the learner's competence level
Output: PRE; POST
begin
    PRE := OR;
    switch Competence do
        case very low do
            if Emotional stability = low then
                POST := N;
            else
                POST:= 2N;
            end
        end
        case moderate do
            if Emotional stability = low then
                PRE := R;
                POST := N;
            else
                POST := NR;
            end
        end
        case outstanding do
            PRE := 2R;
            if Emotional stability = low then
                POST := N;
            else
                POST := 2N;
            end
        end
    end
end
```

significant effects of ES on POST knowledge, and competence on both PRE and POST knowledge. A further exploratory analysis suggests that selecting LAs with less POST or PRE knowledge is better for low ES learners in terms of suitability and to increase confidence. Based on the data analysis, three algorithms have been constructed to adapt LA selection to different levels of ES and competence (see summary in **Table 7**).

5. EVALUATION AND REFINEMENT OF ALGORITHMS

Above, we created three algorithms to adapt the selection of LAs to learner personality (ES) and competence. This section describes an evaluation of key aspects of these algorithms with teachers, resulting in a final algorithm.

5.1. Participants

Twenty-seven participants took part. Six were excluded from the study due to their incorrect answer to the verification question. The final sample consisted of 21 participants (11 female, 9 male, 1 non-disclosed; 9 26–35, 7 36–45, 3 over 46, and 2 prefer not to say; 10 teachers, 11 trainee-teachers).

5.2. Materials

We used the following materials:

1. Two stories depicting ES at either a low or high level developed by Dennis et al. (2012).
2. Three validated levels of competence: very low, moderate, and outstanding.
3. Seven LAs selected based on the three algorithms produced above (LAs 8–12, 16, 18 from **Figure 1**). LAs were shown as before.

5.3. Procedure

Ethical approval was obtained from the University of Aberdeen's Engineering and Physical Sciences ethics board. Before taking

TABLE 6 | Model predictions for PRE and POST.

Competence	ES	PRE	Calculated value	Median PRE	POST	Calculated value	Median POST
Very low	High	1 (O)	0.107	3	1 (N)	0.391	2
		2 (2O or R)	0.304		2 (NO)	0.261	
		3 (OR)	0.438		3 (NR)	0.145	
		4 (2R)	0.150		4 (2N)	0.202	
	Low	1 (O)	0.125	3	1 (N)	0.635	2
		2 (2O or R)	0.329		2 (NO)	0.200	
		3 (OR)	0.416		3 (NR)	0.079	
		4 (2R)	0.129		4 (2N)	0.086	
Moderate	High	1 (O)	0.064	3	1 (N)	0.188	3
		2 (2O or R)	0.220		2 (NO)	0.215	
		3 (OR)	0.478		3 (NR)	0.184	
		4 (2R)	0.236		4 (2N)	0.412	
	Low	1 (O)	0.075	3	1 (N)	0.386	2
		2 (2O or R)	0.246		2 (NO)	0.261	
		3 (OR)	0.471		3 (NR)	0.147	
		4 (2R)	0.206		4 (2N)	0.206	
Outstanding	High	1 (O)	0.037	3	1 (N)	0.077	3
		2 (2O or R)	0.147		2 (NO)	0.119	
		3 (OR)	0.462		3 (NR)	0.143	
		4 (2R)	0.352		4 (2N)	0.660	
	Low	1 (O)	0.044	3	1 (N)	0.185	3
		2 (2O or R)	0.168		2 (NO)	0.214	
		3 (OR)	0.473		3 (NR)	0.184	
		4 (2R)	0.313		4 (2N)	0.418	

Algorithm 3: LA selection based on the regression analyses

```
Input: Emotional stability the learner's level of emotional
       stability; competence the learner's competence level
Output: PRE; POST
begin
  PRE := OR;
  switch Competence do
    case very low do
      POST := NO;
    end
    case moderate do
      if Emotional stability = low then
        POST := NO;
      else
        POST := NR;
      end
    end
    case outstanding do
      POST := NR;
    end
  end
end
```

TABLE 7 | Predictions of LA selections.

Competence	ES	LAs selection					
		Algorithm 1		Algorithm 2		Algorithm 3	
		PRE	POST	PRE	POST	PRE	POST
Very low	High	OR	NO	OR	2N	OR	NO
	Low	OR	N	OR	N	OR	NO
Moderate	High	OR	N	OR	NR	OR	NR
	Low	OR	N	R	N	OR	NO
Outstanding	High	OR	2N	2R	2N	OR	NR
	Low	OR	2N	2R	N	OR	NR

part, participants provided informed consent. Participants first provided demographic information (age, gender and occupation). They were shown two scenarios, one depicting Josh who was high on ES and another depicting James who was low on ES. They were told that the learners had previously learned topics A and B, and recently finished a learning activity which taught topics D and E. For each scenario, three questions were asked, each highlighting a different competence level (*very low, moderate, outstanding*). Participants ranked a subset of the seven LAs, based on their suitability for that learner. **Table 8** shows for each level of competence and ES which LAs participants ranked, using the PRE and POST to describe the LAs. These LAs were chosen such that they included the LAs recommended by each of the three algorithms for that combination of competence and ES (as denoted in **Table 8**), as well as any LAs recommended by the algorithms for that level of competence but for the opposite ES.

TABLE 8 | Median and average for LAs' rankings.

Competence	Emotional stability	LAs		Proposed by Algorithm	Median	Average	Chosen for Algorithm 4	
		PRE	POST				PRE	POST
Very low	High	OR	N		**2**	1.76	OR	NO
		OR	NO	1, 3	**2**	**1.57**		
		OR	2N	2	3	2.67		
	Low	OR	N	1, 2	**2**	1.62	OR	NO
		OR	NO	3	**2**	**1.57**		
		OR	2N		3	2.81		
Moderate	High	OR	N	1	**2**	2.29	OR	NO
		OR	NO		**2**	**2.19**		
		OR	NR	2, 3	3	2.67		
		R	N		4	2.86		
	Low	OR	N	1	**2**	2.38	OR	NO
		OR	NO	3	**2**	**2.14**		
		OR	NR		3	2.81		
		R	N	2	3	2.67		
Outstanding	High	OR	NR	3	4	3.29	2R	2N
		OR	2N	1	3	2.52		
		2R	N		**2**	2.43		
		2R	2N	2	**2**	**1.76**		
	Low	OR	NR	3	**2**	2.48	2R	N
		OR	2N	1	3	2.95		
		2R	N	2	**2**	**2.05**		
		2R	2N		3	2.52		

Bold in Median means best median, and bold in Average means the best average among the LAs rankings.

5.4. Research Questions

We investigated the following research questions:

1. For each level of learner competence and ES, how highly are the selected LAs by Algorithm 1, Algorithm 2, and Algorithm 3 ranked by the teachers, and which LA is ranked highest?
2. Which algorithm matches the rankings of the teachers best?
3. What modifications are needed to the best algorithm to be in line with teachers' preferences?

5.5. Results

Table 8 and **Figure 7** show the results of the ranking. We calculated both the average rank and the median rank.

5.5.1. Very Low Competence

For high ES, the teachers' ranking is best for OR→NO, in line with the predictions of Algorithms 1 and 3. For low ES, the teachers ranking is also best for OR→NO, matching the prediction of Algorithm 3. The prediction by Algorithm 2 did badly in the high ES case, with teachers clearly preferring less complicated LAs than Algorithm 2 had predicted. In fact LAs that involved more new knowledge to learn (2N) were deemed to be the least suited LAs for both ES levels. OR→NO and OR→N did about equally well in the low ES case, so overall the predictions by Algorithm 1 are also good. The teachers clearly where in two minds on whether adaptation to ES would be a good idea for learners with very low competence. Follow on studies measuring learners' attainment and motivation should show whether it is better to use OR→NO or OR→N for low ES learners.

5.5.2. Moderate Competence

For high ES, the teachers' ranking is best for OR→NO. This is not predicted by any of the algorithms. Algorithm 1 predicted a less complicated LA, namely OR→N, whilst Algorithms 2 and 3 predicted a more complicated LA, namely OR→NR. Teachers went for an LA in between, with the ranking of that LA close to that predicted by Algorithm 1. For low ES, the teachers' ranking is best for OR→NO, in line with the prediction of Algorithm 3. Algorithm 2 did badly for both levels of ES. For moderate competence, there is no evidence of adapting to ES levels.

5.5.3. Outstanding Competence

For high ES, the teachers' ranking is best for 2R→2N, in line with the prediction by Algorithm 2. We recall that two LAs scored equally well when constructing Algorithm 1. We selected OR→2N at the time, given the lack of a statistically significant effect of PRE. The alternative was 2R→2N. The teachers clearly preferred the latter one. For low ES, the teachers' ranking is best for 2R→N, again in line with the prediction of Algorithm 2, and showing that teachers are adapting their rankings based on ES, using less new knowledge to learn for the low ES learner. Overall, for outstanding competence, Algorithm 2's predictions were perfect. For both levels of ES, teachers ranked the two LAs that required only recent knowledge higher

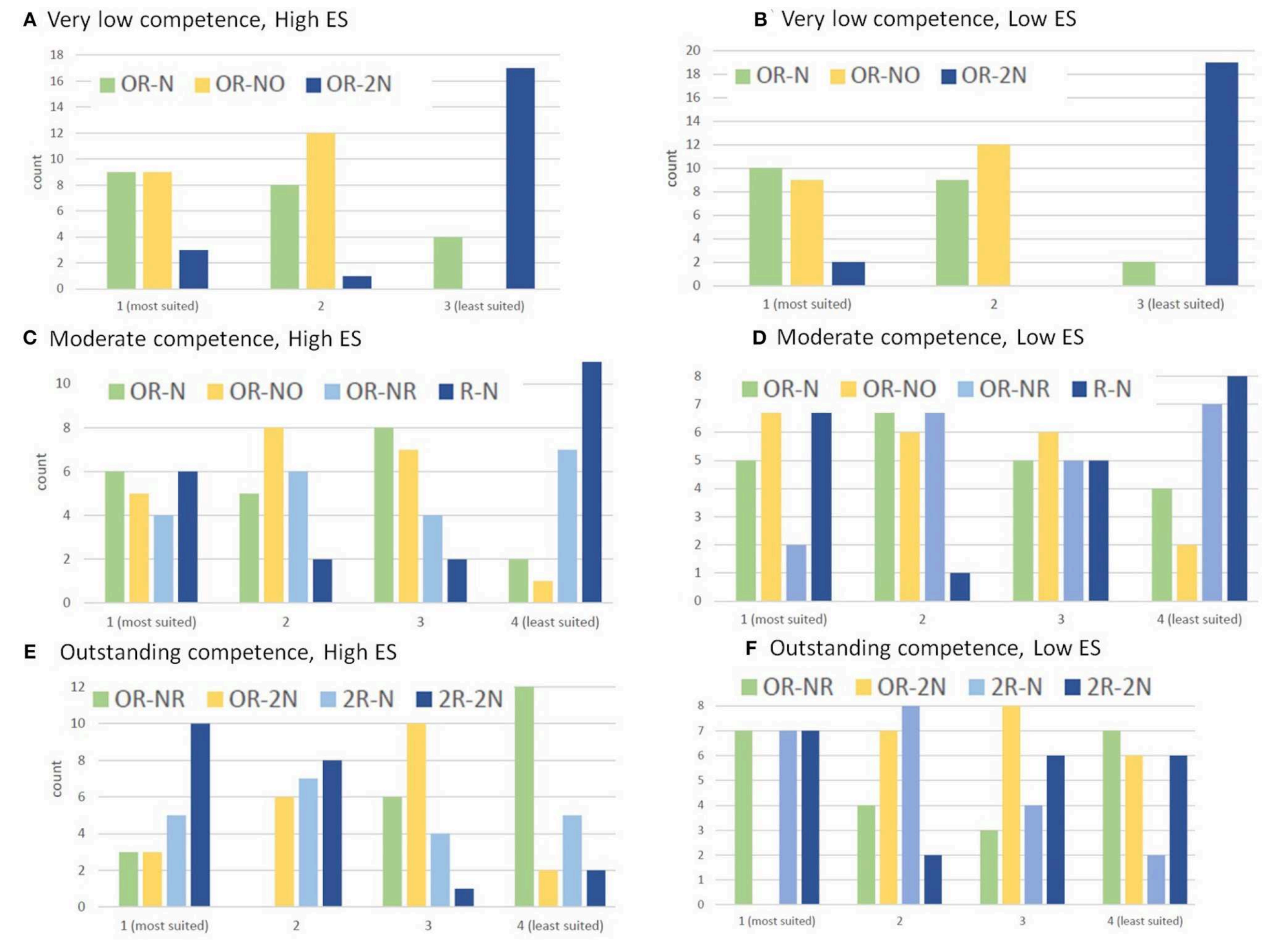


FIGURE 7 | Teachers' rankings for LAs for **(A,B)** very low competence, **(C,D)** moderate competence, and **(E,F)** outstanding competence, for **(A,C,E)** high ES and **(B,D,F)** low ES.

than the two LAs that required a combination of old an recent knowledge, showing an inclination to only use recent knowledge for outstanding competence.

5.6. Refining the Algorithms

We did not find that one algorithm performed better than the others. Algorithm 3 performed best for the Very low competence case (and Algorithm 1 almost equally well), and also for low ES in the Moderate competence case. In contrast, Algorithm 2 performed best for the Outstanding competence case, but badly in the other ones. We decided to produce a new algorithm, combing elements from Algorithms 3 and 2. **Table 8** shows the selections of LAs made for Algorithm 4, which were based on the best median rankings by the teachers. The resulting algorithm is shown in Algorithm 4.

6. CONCLUSION

This paper investigated the impact of learner personality (emotional stability) and competence on the selection of a LA based on the knowledge it uses and the knowledge it teaches. We also investigated the extent to which the selected LAs are perceived to be enjoyable and to increase learners' confidence and skills. ES and competence both impacted the selection of LAs. There were significant effects of ES on POST knowledge, and competence on both PRE and POST knowledge. A further exploratory analysis suggests that selecting LAs with less POST or PRE knowledge is better for low ES learners in terms of suitability and to increase confidence.

Based on the data analysis, an algorithm has been constructed to adapt LA selection to different levels of ES and competence. we obtained four algorithms for adapting LA selection to learners' personality and competence. Algorithms 3 and 4 are the most promising to investigate further, with Algorithm 4 best matching the teachers' preferences, and Algorithm 3 being most aligned to the teachers' preferences from the algorithms based on the data in study 4. These algorithms can be used in an Intelligent Tutoring System, or, as we recommend in future work, can be used as a basis for further research. In addition, we obtained an insight into how teachers adapt LA selection and how this matches

Algorithm 4: LA selection based on teachers' preferences

```
Input: Emotional stability the learner's level of emotional
       stability; competence the learner's competence level
Output: PRE; POST
begin
    PRE := OR;
    switch Competence do
        case very low do
            POST := NO;
        end
        case moderate do
            POST := NO;
        end
        case outstanding do
            PRE := 2R;
            if Emotional stability = low then
                POST := N;
            else
                POST := 2N;
            end
        end
    end
end
```

the algorithms developed. We found evidence that teachers take emotional stability into account when selecting different LAs.

This paper has several limitations and opportunities for future work. First, we did not measure *actual* enjoyment, increase in confidence and increase in skills, but perceptions of those. Studies with learners and real learning tasks are needed to investigate actual impact. Second, the studies in this paper used an abstract notation for learning topics, using letters, such as A, E to indicate which concepts are needed to be known to study something, and which concepts are learned in an activity. This was done on purpose, so that we could study learning activity selection without participants' preconceived ideas about difficulty level of individual concepts and learning domains interfering. However, clearly further studies need to show to what extent what was learned in this paper can be generalized to real learning topics. Further studies are also needed to investigate the possible impacts of learning domains. Third, our algorithm requires a certain structure of the learning activities, namely what is taught (i.e., learning outcomes) and what is used (i.e., prerequisites) in a learning activity. It also requires a learner model in terms of these outcomes, so that we know what a learner has already studied. This may limit its applicability, however, the use of learning outcomes (and also prerequisites) is well-established, and strongly advocated in educational science (Kennedy, 2006). Fourth, we only investigated three levels of competence and ES only at the high and low level. The competence level validation reported in this paper would allow investigating another two levels. It would also be interesting to investigate finer gradations of ES. Fifth, other learner characteristics could be investigated, for example, the impact of learner goals and interests, or as advocated by Zhu et al. (2019) participation levels. As initial research by Adamu Sidi-Ali et al. (2019) showed that cultural background may impact desired learner emotional support, we would also like to investigate whether cultural background should matter for learning activity selection. Sixth, we did not consider other personality traits. Based on previous research (Okpo et al., 2018), we expect learner self-esteem to also matter. Seventh, this paper does not consider how long ago previous topics were studied. A forgetting model will be needed to take into account the likelihood that a learner still masters a topic or that a topic may need to be used in order to prevent forgetting (see Ilbeygi et al., 2019 for an overview and recent work on forgetting models). Eight, this paper only considered learning activity selection for individual learners. This becomes an even more complicated issue when learning activities need to be selected for groups of learners for a collaborative learning experience. Finally, we only considered PRE and POST knowledge, but did not explicitly address difficulty levels.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by a University of Aberdeen ethics committee. The participants provided their written informed consent to participate in the studies.

AUTHOR CONTRIBUTIONS

MA and JM designed and analyzed the studies, and wrote the paper. NB contributed to the study design and helped to improve the paper.

ACKNOWLEDGMENTS

The Ph.D. of MA has been supported by Princess Nourah Bint Abdul Rahman University in Saudi Arabia.

REFERENCES

Adamu Sidi-Ali, M., Masthoff, J., and Dennis, M. (2019). "Adapting performance and emotional support feedback to cultural differences," in *27th ACM Conference on User Modeling, Adaptation and Personalisation, UMAP 2019* (Larnaca). doi: 10.1145/3320435.3320444

Alhathli, M., Masthoff, J., and Siddharthan, A. (2016). "Exploring the impact of extroversion on the selection of learning materials," in *Proceedings of the 6th Workshop on Personalization Approaches in Learning Environments at UMAP. CEUR Workshop Proceedings* (Halifax, NS).

Alhathli, M., Masthoff, J., and Siddharthan, A. (2017). "Should learning material's selection be adapted to learning style and personality?" in *Adjunct Publication* *of the 25th Conference on User Modeling, Adaptation and Personalization* (Bratislava: ACM).

Alhathli, M. A. E., Masthoff, J. F. M., and Beacham, N. A. (2018a). "The effects of learners' verbal and visual cognitive styles on instruction selection," in *Proceedings of Intelligent Mentoring Systems Workshop Associated With the 19th International Conference on Artificial Intelligence in Education, AIED 2018* (London).

Alhathli, M. A. E., Masthoff, J. F. M., and Beacham, N. A. (2018b). "Impact of a learner's personality on the selection of the next learning activity," in *Proceedings of Intelligent Mentoring Systems Workshop Associated With the 19th International Conference on Artificial Intelligence in Education, AIED 2018* (London).

Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: lessons learned. *J. Learn. Sci.* 4, 167–207. doi: 10.1207/s15327809jls0402_2

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*. White Plains, NY: Longman.

Anderson, M., Ball, M., Boley, H., Greene, S., Howse, N., Lemire, D., et al. (2003). "Racofi: A Rule-Applying Collaborative Filtering System," in *Proceedings of IEEE/WIC COLA'03* (Halifax, NS).

Avancini, H., and Straccia, U. (2005). User recommendation for collaborative and personalised digital archives. *Int. J. Web Based Commun.* 1, 163–175. doi: 10.1504/IJWBC.2005.006061

Ayersman, D. J., and von Minden, A. (1995). Individual differences, computers, and instruction. *Comput. Hum. Behav.* 11, 371–390. doi: 10.1016/0747-5632(95)80005-S

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educ. Psychol.* 40, 199–209. doi: 10.1207/s15326985ep4004_2

Barrick, M. R. (2005). Yes, personality matters: moving on to more important matters. *Hum. Perform.* 18, 359–372. doi: 10.1207/s15327043hup1804_3

Barrick, M. R., Mount, M. K., and Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Int. J. Select. Assess.* 9, 9–30. doi: 10.1111/1468-2389.00160

Beal, C., and Lee, H. (2005a). "Creating a pedagogical model that uses student self reports of motivation and mood to adapt its instruction," in *Workshop on Motivation and Affect in Educational Software, in Conjunction With the 12th International Conference on Artificial Intelligence in Education*, Vol. 574 (Amsterdam).

Beal, C. R., and Lee, H. (2005b). "Creating a pedagogical model that uses student self reports of motivation and mood to adapt its instruction," in *Proceedings of the Workshop on Motivation and Affect in Educational Software* (Amsterdam).

Beaumont, I., and Brusilovsky, P. (1995). "Educational applications of adaptive hypermedia," in *Human–Computer Interaction*, eds K. Nordby, P. Helmersen, D. J. Gilmore, and S. A. Arnesen (Boston, MA: Springer), 410–414. doi: 10.1007/978-1-5041-2896-4_71

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Book 1: Cognitive Domain.* New York, NY: Longman.

Bong, M. (2001). Between-and within-domain relations of academic motivation among middle and high school students: self-efficacy, task value, and achievement goals. *J. Educ. Psychol.* 93:23. doi: 10.1037/0022-0663.93.1.23

Bozionelos, N. (2004). The relationship between disposition and career success: a British study. *J. Occupat. Organ. Psychol.* 77, 403–420. doi: 10.1348/0963179041752682

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Model. User Adapt. Interact.* 6, 87–129. doi: 10.1007/BF00143964

Brusilovsky, P. (1998a). "Methods and techniques of adaptive hypermedia," in *Adaptive Hypertext and Hypermedia*, eds P. Brusilovsky, A. Kobsa, and J. Vassileva (Dordrecht: Springer), 1–43. doi: 10.1007/978-94-017-0617-9_1

Brusilovsky, P. (1998b). "Adaptive educational systems on the world-wide-web: a review of available technologies," in *Proceedings of Workshop "WWW-Based Tutoring" at 4th International Conference on Intelligent Tutoring Systems (ITS'98)* (San Antonio, TX).

Brusilovsky, P. (2001). Adaptive hypermedia: methods and techniques. *Int. J. User Model. User Adapt. Interact.* 11, 87–110. doi: 10.1023/A:1011143116306

Brusilovsky, P. (2003). Adaptive navigation support in educational hypermedia: the role of student knowledge level and the case for meta-adaptation. *Br. J. Educ. Technol.* 34, 487–497. doi: 10.1111/1467-8535.00345

Brusilovsky, P. (2007). "Adaptive navigation support," in *The Adaptive Web*, eds P. Brusilovsky, A. Kobsa, and W. Nejdl (Berlin; Heidelberg: Springer), 263–290. doi: 10.1007/978-3-540-72079-9_8

Brusilovsky, P., Eklund, J., and Schwarz, E. (1998). Web-based education for all: a tool for development adaptive courseware. *Comput. Netw. ISDN Syst.* 30, 291–300. doi: 10.1016/S0169-7552(98)00082-8

Brusilovsky, P., and Millán, E. (2007). "User models for adaptive hypermedia and adaptive educational systems," in *The Adaptive Web*, eds P. Brusilovsky, A. Kobsa, and W. Nejdl (Berlin; Heidelberg: Springer), 3–53. doi: 10.1007/978-3-540-72079-9_1

Brusilovsky, P., Schwarz, E., and Weber, G. (1996). "Elm-art: an intelligent tutoring system on world wide web," in *International Conference on Intelligent Tutoring Systems* (Montreal, QC: Springer), 261–269. doi: 10.1007/3-540-61327-7_123

Bunderson, C. V., and Martinez, M. (2000). Building interactive world wide web (web) learning environments to match and support individual learning differences. *J. Interact. Learn. Res.* 11, 163–195.

Butler, K. A. (1988). How kids learn: what theorists say. *Learning* 17, 30–34.

Byrnes, J. P., and Guthrie, J. T. (1992). Prior conceptual knowledge and textbook search. *Contemp. Educ. Psychol.* 17, 8–29. doi: 10.1016/0361-476X(92)90042-W

Camp, G., Paas, F., Rikers, R., and van Merrienboer, J. (2001). Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency. *Comput. Hum. Behav.* 17, 575–595. doi: 10.1016/S0747-5632(01)00028-0

Chang, H., Cohn, D., and Mccallum, A. K. (2000). "Learning to create customized authority lists," in *ICML'00: Proceedings of the Seventeenth International Conference on Machine Learning* (Stanford, CA), 127–134.

Cheng, I., Shen, R., and Basu, A. (2008). "An algorithm for automatic difficulty level estimation of multimedia mathematical test items," in *Advanced Learning Technologies, 2008. ICALT'08* (Santander: IEEE), 175–179. doi: 10.1109/ICALT.2008.105

Clark, M., and Schroth, C. A. (2010). Examining relationships between academic motivation and personality among college students. *Learn. Individ. Differ.* 20, 19–24. doi: 10.1016/j.lindif.2009.10.002

Cocea, M., and Weibelzahl, S. (2006). "Motivation: included or excluded from e-learning," in *Cognition and Exploratory Learning in Digital Age, CELDA 2006 Proceedings* (Barcelona: IADIS Press).

Corbalan, G., Kester, L., and Van Merriënboer, J. J. (2008). Selecting learning tasks: effects of adaptation and shared control on learning efficiency and task involvement. *Contemp. Educ. Psychol.* 33, 733–756. doi: 10.1016/j.cedpsych.2008.02.003

Costa, P., and McCrae, R. (1992a). *Neo PI-R: Professional Manual: Revised Neo PI-R and Neo-FFI*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., and McCrae, R. R. (1992b). Four ways five factors are basic. *Pers. Individ. Differ.* 13, 653–665. doi: 10.1016/0191-8869(92)90236-I

Costa, P. T., and McCrae, R. R. (1995). Solid ground in the wetlands of personality: a reply to block. *Psychol. Bull.* 117, 216–220; discussion 226–229. doi: 10.1037/0033-2909.117.2.216

Cristea, A. I. (2003). Adaptive patterns in authoring of educational adaptive hypermedia. *Educ. Technol. Soc.* 6, 1–5. doi: 10.2316/Journal.208.2004.4.208-0826

Dagger, D., Wade, V., and Conlan, O. (2005). Personalisation for all: making adaptive course composition easy. *J. Educ. Technol. Soc.* 8, 9–25.

Dascalu, M.-I., Bodea, C.-N., Mihailescu, M. N., Tanase, E. A., and Ordoñez de Pablos, P. (2016). Educational recommender systems and their application in lifelong learning. *Behav. Inform. Technol.* 35, 290–297. doi: 10.1080/0144929X.2015.1128977

Davidovic, A., Warren, J., and Trichina, E. (2003). Learning benefits of structural example-based adaptive tutoring systems. *IEEE Trans. Educ.* 46, 241–251. doi: 10.1109/TE.2002.808240

De Bra, P., Aroyo, L., and Cristea, A. (2004). "Adaptive web-based educational hypermedia," in *Web Dynamics*, eds M. Levene and A. Poulovassilis (Berlin; Heidelberg: Springer), 387–410. doi: 10.1007/978-3-662-10874-1_16

Dennis, M., Masthoff, J., and Mellish, C. (2012). "The quest for validated personality trait stories," in *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces* (Lisbon: ACM), 273–276. doi: 10.1145/2166966.2167016

Dennis, M., Masthoff, J., and Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *Int. J. Artif. Intell. Educ*. 26, 877–931. doi: 10.1007/s40593-015-0059-7

Di Iorio, A., Feliziani, A. A., Mirri, S., Salomoni, P., and Vitali, F. (2006). Automatically producing accessible learning objects. *Educ. Technol. Soc*. 9, 3–16.

Dochy, F. (1994). "Prior knowledge and learning," in *International Encyclopedia of Education, 2nd Edn*, eds T. Husen and T. N. Postlethwaite (Oxford; New York, NY: Pergamon Press), 4698–4702.

Dochy, F. J. R. C. (1992). *Assessment of Prior Knowledge as a Determinant for Future Learning: The Use of Prior Knowledge State Tests and Knowledge Profiles*. Milton Keynes: Centre for Educational Technology and Innovation, Open University.

Drachsler, H., Hummel, H., and Koper, R. (2008a). Identifying the goal, user model and conditions of recommender systems for formal and informal learning. *J. Digit. Inf.* 10, 4–24.

Drachsler, H., Hummel, H. G., and Koper, R. (2008b). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *Int. J. Learn. Technol*. 3, 404–423. doi: 10.1504/IJLT.2008.019376

El-Bishouty, M. M., Chang, T.-W., Graf, S., and Chen, N.-S. (2014). Smart e-course recommender based on learning styles. *J. Comput. Educ*. 1, 99–111. doi: 10.1007/s40692-014-0003-0

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educ. Psychol*. 34, 169–189. doi: 10.1207/s15326985ep3403_3

Elliot, A. J., and McGregor, H. A. (2001). A 2× 2 achievement goal framework. *J. Pers. Soc. Psychol*. 80:501. doi: 10.1037/0022-3514.80.3.501

Elliot, A. J., and Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educ. Psychol. Rev*. 13, 139–156. doi: 10.1023/A:1009057102306

Forbes-Riley, K., Rotaru, M., and Litman, D. J. (2008). The relative impact of student affect on performance models in a spoken dialogue tutoring system. *User Model. User Adapt. Interact*. 18, 11–43. doi: 10.1007/s11257-007-9038-5

Ford, N., and Chen, S. Y. (2000). Individual differences, hypermedia navigation, and learning: an empirical study. *J. Educ. Multimed. Hypermed*. 9, 281–311.

Gaultney, J. F. (1995). The effect of prior knowledge and metacognition on the acquisition of a reading comprehension strategy. *J. Exp. Child Psychol*. 59, 142–163. doi: 10.1006/jecp.1995.1006

Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., and Samaras, G. (2008). Capturing essential intrinsic user behaviour values for the design of comprehensive web-based personalized environments. *Comput. Hum. Behav*. 24, 1434–1451. doi: 10.1016/j.chb.2007.07.010

Gilbert, J. E., and Han, C. Y. (1999). Adapting instruction in search of 'a significant difference'. *J. Netw. Comput. Appl*. 22, 149–160. doi: 10.1006/jnca.1999.0088

Graesser, A. C., D'Mello, S. K., Craig, S. D., Witherspoon, A., Sullins, J., McDaniel, B., et al. (2008). The relationship between affective states and dialog patterns during interactions with autotutor. *J. Interact. Learn. Res*. 19:293–312.

Graesser, A. C., Jackson, G. T., and McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educ. Technol.* 47, 19–23.

Hazrati-Viari, A., Rad, A. T., and Torabi, S. S. (2012). The effect of personality traits on academic performance: the mediating role of academic motivation. *Proc. Soc. Behav. Sci*. 32, 367–371. doi: 10.1016/j.sbspro.2012.01.055

Henze, N., Naceur, K., Nejdl, W., and Wolpers, M. (1999). Adaptive hyperbooks for constructivist teaching. *KI* 13, 26–31.

Hospers, M., Kroezen, E., Nijholt, A., op den Akker, R., and Heylen, D. (2003). "An agent-based intelligent tutoring system for nurse education," in *Applications of Software Agent Technology in the Health Care Domain*, eds A. Moreno and J. L. Nealon (Basel: Springer), 143–159. doi: 10.1007/978-3-0348-7976-7_10

Hsu, M.-H. (2008). A personalized english learning recommender system for esl students. *Exp. Syst. Appl*. 34, 683–688. doi: 10.1016/j.eswa.2006.10.004

Huang, Y.-M., Huang, T.-C., Wang, K.-T., and Hwang, W.-Y. (2009). A markov-based recommendation model for exploring the transfer of learning on the web. *J. Educ. Technol. Soc*. 12, 144–162.

Hwang, G.-J. (1998). A tutoring strategy supporting system for distance learning on computer networks. *IEEE* 41, 343–343. doi: 10.1109/TE.1998.787369

Ilbeygi, M., Kangavari, M. R., and Golmohammadi, S. A. (2019). Equipping the act-r cognitive architecture with a temporal ratio model of memory and using it in a new intelligent adaptive interface. *User Model. User Adapt. Interact*. 29, 943–976. doi: 10.1007/s11257-019-09239-2

John, O. P., and Srivastava, S. (1999). The big five trait taxonomy: history, measurement, and theoretical perspectives. *Handb. Pers. Theory Res*. 2, 102–138.

Jonassen, D., and Grabowski, B. (1993). *Individual Differences and Instruction*. New York, NY: Allen & Bacon.

Judge, T. A., Higgins, C. A., Thoresen, C. J., and Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Pers. Psychol*. 52, 621–652. doi: 10.1111/j.1744-6570.1999.tb00174.x

Kalyuga, S., and Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educ. Technol. Res. Dev*. 53, 83–93. doi: 10.1007/BF02504800

Kelly, D. (2008). Adaptive versus learner control in a multiple intelligence learning environment. *J. Educ. Multimed. Hypermed*. 17, 307–336.

Kelly, D., and Tangney, B. (2002). "Incorporating learning characteristics into an intelligent tutor," in *International Conference on Intelligent Tutoring Systems* (Biarritz: Springer), 729–738. doi: 10.1007/3-540-47987-2_73

Kennedy, D. (2006). *Writing and Using Learning Outcomes: A Practical Guide*. Cork: University College Cork.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Comput. Educ.* 106, 166–171. doi: 10.1016/j.compedu.2016.12.006

Klemp, G. O. Jr. (1979). Identifying, measuring, and integrating competence. *New Direct. Exp. Learn*. 3, 41–52.

Komarraju, M., and Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Pers. Individ. Differ*. 39, 557–567. doi: 10.1016/j.paid.2005.02.013

Komarraju, M., Karau, S. J., Schmeck, R. R., and Avdic, A. (2011). The big five personality traits, learning styles, and academic achievement. *Pers. Individ. Differ*. 51, 472–477. doi: 10.1016/j.paid.2011.04.019

Koutsojannis, C., Prentzas, J., and Hatzilygeroudis, I. (2001). "A web-based intelligent tutoring system teaching nursing students fundamental aspects of biomedical technology," in *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, Vol. 4 (Istanbul: IEEE), 4024–4027. doi: 10.21236/ADA412356

Latham, A., Crockett, K., McLean, D., and Edmonds, B. (2012). A conversational intelligent tutoring system to automatically predict learning styles. *Comput. Educ*. 59, 95–109. doi: 10.1016/j.compedu.2011.11.001

Lawless, K. A., and Brown, S. W. (1997). Multimedia learning environments: issues of learner control and navigation. *Instruct. Sci*. 25, 117–131. doi: 10.1023/A:1002919531780

Lee, J., and Park, O. (2008). "Adaptive instructional systems," in *Handbook of Research on Educational Communications and Technology*, eds J. M Spector, M. D. Merrill, J. van Merrienboer, and M. P. Driscoll (New York, NY: Lawrence Erlbaum), 469–484.

Liang, G., Weining, K., and Junzhou, L. (2006). "Courseware recommendation in e-learning system," in *International Conference on Web-Based Learning* (Penang), 10–24. doi: 10.1007/11925293_2

Lo, J.-J., Chan, Y.-C., and Yeh, S.-W. (2012). Designing an adaptive web-based learning system based on students' cognitive styles identified online. *Comput. Educ*. 58, 209–222. doi: 10.1016/j.compedu.2011.08.018

Magoulas, G. D., Papanikolaou, Y., and Grigoriadou, M. (2003). Adaptive web-based learning: accommodating individual differences through system's adaptation. *Br. J. Educ. Technol*. 34, 511–527. doi: 10.1111/1467-8535.00347

Mampadi, F., Chen, S. Y., Ghinea, G., and Chen, M.-P. (2011). Design of adaptive hypermedia learning systems: a cognitive style approach. *Comput. Educ*. 56, 1003–1011. doi: 10.1016/j.compedu.2010.11.018

Masthoff, J. (1997). *An Agent-Based Interactive Instruction System*. Eindhoven: Technical University.

McCrae, R. (1992). The five-factor model: Issues and applications. *J. Pers*. 60, 175–215. doi: 10.1111/j.1467-6494.1992.tb00970.x

McCrae, R. R., and Costa, P. T. Jr. (1997). Personality trait structure as a human universal. *Am. Psychol.* 52:509. doi: 10.1037/0003-066X.52.5.509

Mcquiggan, S. W., Mott, B. W., and Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: an inductive approach. *User Model. User Adapt. Interact.* 18, 81–123. doi: 10.1007/s11257-007-9040-y

Melis, E., Andres, E., Budenbender, J., Frischauf, A., Goduadze, G., Libbrecht, P., et al. (2001). Activemath: a generic and adaptive web-based learning environment. *Int. J. Artif. Intell. Educ.* 12, 385–407.

Miller, A. (1991). Personality types, learning styles and educational goals. *Educ. Psychol.* 11, 217–238. doi: 10.1080/0144341910110302

Mitrović, A., Djordjević-Kajan, S., and Stomenov, L. (1996). Instruct: Modeling students by asking questions. *User Model. User Adapt. Interact.* 6, 273–302. doi: 10.1007/BF00213185

Mitrovic, A., Martin, B., and Mayo, M. (2002). Using evaluation to shape its design: results and experiences with SQL-tutor. *User Model. User Adapt. Interact.* 12, 243–279. doi: 10.1023/A:1015022619307

Nunes, M. A. S., and Hu, R. (2012). "Personality-based recommender systems: an overview," in *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin: ACM), 5–6. doi: 10.1145/2365952.2365957

Oakes, D. W., Ferris, G. R., Martocchio, J. J., Buckley, M. R., and Broach, D. (2001). Cognitive ability and personality predictors of training program skill acquisition and job performance. *J. Bus. Psychol.* 15, 523–548. doi: 10.1023/A:1007805132107

Odo, C. R., Masthoff, J. F. M., Beacham, N. A., and Alhathli, M. A. E. (2018). "Affective state for learning activities selection," in *Proceedings of Intelligent Mentoring Systems Workshop Associated With the 19th International Conference on Artificial Intelligence in Education, AIED 2018* (London). doi: 10.1007/978-3-319-93846-2_98

Okpo, J., Dennis, M., Masthoff, J., Smith, K. A., and Beacham, N. (2016). "Exploring requirements for an adaptive exercise selection system," in *PALE 2016: Workshop on Personalization Approaches in Learning Environments* (Halifax, NS).

Okpo, J., Masthoff, J., Dennis, M., Beacham, N., and Ciocarlan, A. (2017). "Investigating the impact of personality and cognitive efficiency on the selection of exercises for learners," in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17* (New York, NY: ACM), 140–147. doi: 10.1145/3079628.3079674

Okpo, J. A., Masthoff, J., Dennis, M., and Beacham, N. (2018). Adapting exercise selection to performance, effort and self-esteem. *New Rev. Hypermed. Multimed.* 24, 193–227. doi: 10.1080/13614568.2018.1477999

Ones, D. S., Dilchert, S., Viswesvaran, C., and Judge, T. A. (2007). In support of personality assessment in organizational settings. *Pers. Psychol.* 60, 995–1027. doi: 10.1111/j.1744-6570.2007.00099.x

Papanikolaou, K. A., Grigoriadou, M., Kornilakis, H., and Magoulas, G. D. (2001). "INSPIRE: an intelligent system for personalized instruction in a remote environment," in *Workshop on Adaptive Hypermedia.* (Berlin; Heidelberg: Springer), 215–225.

Park, O.-C, and Lee, J. (2004). Adaptive instructional systems. in *Handbook of research for education communications and technology, 2nd Edn*, ed D. H. Jonassen. (Mahwah, NJ: Erlbaum), 651–684.

Parry, S. B. (1996). The quest for competencies. *Training* 33:48.

Petrovica, S. (2013). "Adaptation of tutoring to students' emotions in emotionally intelligent tutoring systems," in *2013 Second International Conference on e-Learning and e-Technologies in Education (ICEEE)* (Lodz: IEEE), 131–136. doi: 10.1109/ICeLeTE.2013.6644361

Pholo, D., and Ngwira, S. (2013). "Integrating explicit problem-solving teaching into activemath, an intelligent tutoring system," in *2013 International Conference on Adaptive Science and Technology (ICAST)* (Pretoria: IEEE), 1–8. doi: 10.1109/ICASTech.2013.6707521

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychol. Bull.* 135:322. doi: 10.1037/a0014996

Prins, F., Nadolski, R., Berlanga, A., Drachsler, H., Hummel, H., and Koper, R. (2007). Competence description for personal recommendations: the importance of identifying the complexity of learning and performance situations. *Educ. Technol. Soc.* 11, 141–152.

Rafaeli, S., Barak, M., Dan-Gur, Y., and Toch, E. (2004). Qsia-a web-based environment for learning, assessing and knowledge sharing in communities. *Comput. Educ.* 43, 273–289. doi: 10.1016/j.compedu.2003.10.008

Ray, R. D., and Belden, N. (2007). Teaching college level content and reading comprehension skills simultaneously via an artificially intelligent adaptive computerized instructional system. *Psychol. Rec.* 57, 201–218. doi: 10.1007/BF03395572

Reategui, E., Boff, E., and Campbell, J. A. (2008). Personalization in an interactive learning environment through a virtual character. *Comput. Educ.* 51, 530–544. doi: 10.1016/j.compedu.2007.05.018

Recker, M. M., and Walker, A. (2003). Supporting "word-of-mouth" social networks through collaborative information filtering. *J. Interact. Learn. Res.* 14, 79–98.

Recker, M. M., Walker, A., and Lawless, K. (2003). What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. *Instruct. Sci.* 31, 299–316. doi: 10.1023/A:1024686010318

Ree, M. J., Earles, J. A., and Teachout, M. S. (1994). Predicting job performance: not much more than G. *J. Appl. Psychol.* 79:518. doi: 10.1037/0021-9010.79.4.518

Revilla, M. A., Manzoor, S., and Liu, R. (2008). Competitive learning in informatics: the UVA online judge experience. *Olymp. Inform.* 2, 131–148.

Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. in *Recommender Systems Handbook*, eds F. Ricci, L. Rokach, B. Shapira, and P. Kantor (Boston, MA: Springer). doi: 10.1007/978-0-387-85820-3_1

Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* 138:353. doi: 10.1037/a0026838

Robison, J., McQuiggan, S., and Lester, J. (2010). "Developing empirically based student personality profiles for affective feedback models," in *International Conference on Intelligent Tutoring Systems* (Pittsburgh, PA: Springer), 285–295. doi: 10.1007/978-3-642-13388-6_33

Rumetshofer, H., and Wöß, W. (2003). Xml-based adaptation framework for psychological-driven e-learning systems. *J. Educ. Technol. Soc.* 6, 18–29.

Salden, R. J., Paas, F., Broers, N. J., and Van Merriënboer, J. J. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instruct. Sci.* 32, 153–172. doi: 10.1023/B:TRUC.0000021814.03996.ff

Salden, R. J., Paas, F., and Van Merriënboer, J. J. (2006). Personalised adaptive task selection in air traffic control: effects on training efficiency and transfer. *Learn. Instruct.* 16, 350–362. doi: 10.1016/j.learninstruc.2006.07.007

Salgado, J. F. (1997). The five factor model of personality and job performance in the european community. *J. Appl. Psychol.* 82:30. doi: 10.1037/0021-9010.82.1.30

Santos, O. C., and Boticario, J. G. (2010). Modeling recommendations for the educational domain. *Proc. Comput. Sci.* 1, 2793–2800. doi: 10.1016/j.procs.2010.08.004

Sawyer, R. K. et al. (2008). Optimising learning implications of learning sciences research. *Innov. Learn Learn Innov.* 45, 35–98. doi: 10.1787/9789264047983-4-en

Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). "Collaborative filtering recommender systems," in *The Adaptive Web*, eds P. Brusilovsky, A. Kobsa, and W. Nejdl (Berlin; Heidelberg: Springer), 291–324. doi: 10.1007/978-3-540-72079-9_9

Schambach, T. P. (1994). *Maintaining Professional Competence: An Evaluation of Factors Affecting Professional Obsolescence of Information Technology Professionals*. PhD Dissertation, University of South Florida, Tampa, FL, United States.

Schiaffino, S., Garcia, P., and Amandi, A. (2008). eteacher: Providing personalized assistance to e-learning students. *Comput. Educ.* 51, 1744–1754. doi: 10.1016/j.compedu.2008.05.008

Schnackenberg, H. L., and Sullivan, H. J. (2000). Learner control over full and lean computer-based instruction under differing ability levels. *Educ. Technol. Res. Dev.* 48, 19–35. doi: 10.1007/BF02313399

Seters, J. v., Ossevoort, M., Goedhart, M., and Tramper, J. (2011). Accommodating the difference in students' prior knowledge of cell growth kinetics. *Electron. J. Biotechnol.* 14:1. doi: 10.2225/vol14-issue2-fulltext-2

Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organ. Sci.* 11, 448–469. doi: 10.1287/orsc.11.4.448.14602

Shen, L.-P., and Shen, R.-M. (2004). *Learning Content Recommendation Service Based-on Simple Sequencing Specification.* Berlin;

Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-27859-7_47
Shute, V., and Towle, B. (2003). Adaptive e-learning. *Educ. Psychol.* 38, 105–114. doi: 10.1207/S15326985EP3802_5
Shute, V. J. (1995). Smart: student modeling approach for responsive tutoring. *User Model. User Adapt. Interact.* 5, 1–44. doi: 10.1007/BF01101800
Shute, V. J., and Zapata-Rivera, D. (2007). Adaptive technologies. *ETS Res. Rep. Ser.* 2007:i-34. doi: 10.1002/j.2333-8504.2007.tb02047.x
Siddappa, M., and Manjunath, A. (2008). "Intelligent tutor generator for intelligent tutoring systems," in *Proceedings of World Congress on Engineering and Computer Science* (San Francisco, CA), 578–583.
Smith, K. A., Dennis, M., Masthoff, J., and Tintarev, N. (2019). A methodology for creating and validating psychological stories for conveying and measuring psychological traits. *User Model. User Adapt. Interact.* 29, 573–618. doi: 10.1007/s11257-019-09219-6
Soonthornphisaj, N., Rojsattarat, E., and Yim-Ngam, S. (2006). "Smart e-learning using recommender system," in *Computational Intelligence*, eds D. S. Huang, K. Li, and G. W. Irwin (Berlin; Heidelberg: Springer), 518–523. doi: 10.1007/978-3-540-37275-2_63
Specht, M., Weber, G., Heitmeyer, S., and Schöch, V. (1997). "AST: adaptive www-courseware for statistics," in *Proceedings of Workshop Adaptive Systems and User Modeling on the World Wide Web at UM97* (Chia Laguna), 91–95.
Stewart, C., Cristea, A. I., Brailsford, T., and Ashman, H. (2005). "'Authoring once, delivering many': creating reusable adaptive courseware," in *International conference on Web-based Education* (Grindelwald).
Sugiyama, K., Hatano, K., and Yoshikawa, M. (2004). "Adaptive web search based on user profile constructed without any effort from users," in *Proceedings of World Wide Web Conference* (ACM), 675–684. doi: 10.1145/988672.988764
Sun, P.-C., and Cheng, H. K. (2007). The design of instructional multimedia in e-learning: a media richness theory-based approach. *Comput. Educ.* 49, 662–676. doi: 10.1016/j.compedu.2005.11.016
Taylor, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *J. Bull.* 30, 415–433. doi: 10.1177/107769905303000401
Thompson, R. A., and Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teach. Psychol.* 30, 96–101. doi: 10.1207/S15328023TOP3002_02
Triantafillou, E., Pomportsis, A., Demetriadis, S., and Georgiadou, E. (2004). The value of adaptivity based on cognitive style: an empirical study. *Br. J. Educ. Technol.* 35, 95–106. doi: 10.1111/j.1467-8535.2004.00371.x
Trotman, A., and Handley, C. (2008). Programming contest strategy. *Comput. Educ.* 50, 821–837. doi: 10.1016/j.compedu.2006.08.008
Tseng, J. C., Chu, H.-C., Hwang, G.-J., and Tsai, C.-C. (2008). Development of an adaptive learning system with two sources of personalization information. *Comput. Educ.* 51, 776–786. doi: 10.1016/j.compedu.2007.08.002
Tsiriga, V., and Virvou, M. (2004). Evaluating the intelligent features of a web-based intelligent computer assisted language learning system. *Int. J. Artif. Intell. Tools* 13, 411–425. doi: 10.1142/S0218213004001600
Vandewaetere, M., Desmet, P., and Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Comput. Hum. Behav.* 27, 118–130. doi: 10.1016/j.chb.2010.07.038
Vassileva, D., and Bontchev, B. (2006). "Self Adaptive Hypermedia Navigation Based on Learner Model Characters," in *Proceedings of IADAT-e2006* (Barcelona), 46–52.
Vassileva, J. (1997). Dynamic course generation. *J. Comput. Inform. Technol.* 5, 87–102.
Verdú, E., Regueras, L. M., Verdú, M. J., Leal, J. P., de Castro, J. P., and Queirós, R. (2012). A distributed system for learning programming on-line. *Comput. Educ.* 58, 1–10. doi: 10.1016/j.compedu.2011.08.015
Wang, M., Hauze, S., Olmstead, W., Aziz, B., Zaineb, B., and Ng, J. (2014). "An exploration of intelligent learning (ilearning) systems," in *IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (Wellington), 327–332. doi: 10.1109/TALE.2014.7062558
Wang, P.-Y., and Yang, H.-C. (2012). Using collaborative filtering to support college students' use of online forum for english learning. *Comput. Educ.* 59, 628–637. doi: 10.1016/j.compedu.2012.02.007
Widyantoro, D. H., Ioerger, T. R., and Yen, J. (1999). "An adaptive algorithm for learning changes in user interests," in *Proceedings of the Eighth International Conference on Information and Knowledge Management* (Kansas City, MO: ACM), 405–412. doi: 10.1145/319950.323230
Willis, S., and Dubin, S. (1990). *Competence Versus Obsolescence: Understanding the Challenge Facing Today's Professionals. Maintaining Professional Competence: Approaches to Career Enhancement, Vitality, and Success Throughout a Work Life*. San Francisco, CA: Josey-Bass.
Wolf, C. (2003). "Iweaver: Towards' learning style'-based e-learning in computer science education," in *Proceedings of the Fifth Australasian Conference on Computing Education*, Vol. 20 (Adelaide, SA: Australian Computer Society, Inc.), 273–279.
Woodruffe, C. (1993). What is meant by a competency? *Leadership* & *organization development journal* 14, 29–36. doi: 10.1108/eb053651
Wünsche, B. C., Chen, Z., Shaw, L., Suselo, T., Leung, K.-C., Dimalen, D., et al. (2018). "Automatic assessment of opengl computer graphics assignments," in *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca: ACM), 81–86. doi: 10.1145/3197091.3197112
Yang, Y. J., and Wu, C. (2009). An attribute-based ant colony system for adaptive learning object recommendation. *Expert Syst. Appl.* 36, 3034–3047. doi: 10.1016/j.eswa.2008.01.066
Zhu, G., Xing, W., Costa, S., Scardamalia, M., and Pei, B. (2019). Exploring emotional and cognitive dynamics of knowledge building in grades 1 and 2. *User Model. User Adapt. Interact.* 29, 789–820. doi: 10.1007/s11257-019-09241-8
Zimmerman, B. J. (2002). "Achieving academic excellence: a self-regulatory perspective," in *The Pursuit of Excellence Through Education*, ed M. Ferrari (New York, NY; Hove: Psychology Press, Taylor & Francis Group), 85–110.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

BRIEF RESEARCH REPORT
published: 09 April 2020
doi: 10.3389/frai.2020.00020

Evaluating Personalization: The AB Testing Pitfalls Companies Might Not Be Aware of—A Spotlight on the Automotive Sector Websites

Maria Esteller-Cucala[1,2]*, *Vicenc Fernandez*[3] *and Diego Villuendas*[2]

[1] Universitat Politècnica de Catalunya-Barcelona Tech., Barcelona, Spain, [2] SEAT, S.A., Barcelona, Spain, [3] TechTalent-Lab, Department of Management, Universitat Politècnica de Catalunya-Barcelona Tech., Barcelona, Spain

The importance of companies' website as instrument for relationship marketing activities is well-known both in the academia and in the industry. In the last decades, there has been great interest in studying how technology can be used to influence people's attitudes and motivate behavior change. With this, web personalization has had increasing research and practitioner interest. However, the evaluation of user interaction with companies' websites and personalization effects remains an elusive goal for organizations. Online controlled experiments (A/B tests) are one of the most commonly known and used techniques for this online evaluation. And, while there is clearly value in evaluating personalized features by means of online controlled experiments, there are some pitfalls to bear in mind while testing. In this paper we present five experimentation pitfalls, firstly identified in an automotive company's website and found to be present in other sectors, that are particularly important or likely to appear when evaluating personalization features. In order to obtain the listed pitfalls, different methods have been used, including literature review, direct, and indirect observation within organizations of the automotive sector and a set of interviews to organizations form other sectors. Finally, the list of five resulting pitfalls is presented and some suggestions are made on how to avoid or mitigate each of them.

Keywords: controlled experiments, online experiments, A/B testing, personalization, online personalization

OPEN ACCESS

Edited by:
Panagiotis Germanakos, SAP SE, Germany

Reviewed by:
Michael Felderer, University of Innsbruck, Austria
Joshua Scotton, Independent Researcher, London, United Kingdom

Correspondence:
Maria Esteller-Cucala
maria.esteller.cucala@gmail.com

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *14 October 2019*
Accepted: *18 March 2020*
Published: *09 April 2020*

Citation:
Esteller-Cucala M, Fernandez V and Villuendas D (2020) Evaluating Personalization: The AB Testing Pitfalls Companies Might Not Be Aware of—A Spotlight on the Automotive Sector Websites. Front. Artif. Intell. 3:20. doi: 10.3389/frai.2020.00020

1. INTRODUCTION

The importance of companies' website as instrument for relationship marketing activities is well-known both in the academia and in the industry (Mahmoud et al., 2017). In the last decades, there has been great interest in studying how technology can be used to influences people's attitudes and motivate behavior change (Oinas-Kukkonen and Harjumaa, 2008b). Moreover, users are nowadays more likely to look for an emotional connection with the interfaces they come across with (Mendoza and Marasinghe, 2013). Accordingly, companies are not anymore using their websites only to inform about their products or services and sell them, they now need to persuade their users to engage with them (Rashid et al., 2016). With this, the evaluation of user interaction with companies websites is in the spotlight (Spiliopoulou, 2000; Yen et al., 2007).

Regardless of the organization size, website owners try to increase users' interface persuasiveness by adapting colors, texts, or layout (Hohnhold et al., 2015). Following this attempt to be continuously improving, the positive effects of website personalization in company pervasiveness have been gaining attention (Kaptein et al., 2015). Web personalization has been proven not only to have a direct effect on user persuasion (Tam et al., 2005; Oinas-Kukkonen and Harjumaa, 2008a), but also to reduce user reference uncertainty and user obfuscation due to information overload (Arora et al., 2008; Xu et al., 2014; Choi et al., 2017). Moreover, it has been proven to increase trustworthiness perception of the organization, satisfaction, user engagement and user loyalty (indirectly by increasing satisfaction and engagement) (Lee and Lin, 2005; Coelho and Henseler, 2012; Xu et al., 2014; Demangeot and Broderick, 2016; Bleier et al., 2017; Piccoli et al., 2017).

This, among other reasons, has set excellent conditions for web personalization to prosper (Salonen and Karjaluoto, 2016). However, while general personalization effects have been proven by the academia, website personalization is a broad concept and determining the specific impact of particular personalized features on an organization's website remains an elusive goal (Kwon et al., 2010; Kaptein et al., 2015). According to that, recent survey results report that "marketers are more unsatisfied with their current efforts and are less confident in their ability to achieve successful personalization" (Researchscape International and Evergage, Inc.). Therefore, there is still not a consensus on how to measure the persuasive effect of personalization (Kaptein and Parvinen, 2015).

Separately, online controlled experiments (also known as A/B tests) play nowadays a significant role in evaluating the impact that website changes have on users (Das and Ranganath, 2013) being one of the most common methods used (Dmitriev et al., 2016). These two trends, accompanied by its simplicity (Knijnenburg, 2012; Bakshy et al., 2014), have created an increasing use of A/B testing to evaluate personalization features on websites (Amatriain and Basilico, 2012; Dmitriev et al., 2017). Evaluating of personalization improvements has become a popular applications in A/B testing (Fabijan et al., 2016; Govind, 2017; Letham et al., 2018). In the simplest case, the experiment participants of an A/B test are randomly split into either one of two comparable groups. The only difference between the groups is some change or variation X deliberately included by the experimenter (from simple changes to personalization algorithms and recommender systems). If the experiment is designed and executed correctly, external factors are distributed evenly between the two groups. Thus, the only thing consistently different between the variants is the change X. Hence, any difference in metrics between the two groups must be due to the change X or a random change (the second being ruled out using statistical testing). Thereby establishing a causal relationship between X and the measured difference in metrics between the two variants (Kohavi et al., 2007; Crook et al., 2009; Fabijan et al., 2016; Zhao and Zhao, 2016; Johari et al., 2017). With the rise of software and internet connectivity, A/B testing presents an unprecedented opportunity to make causal conclusions between the changes made and the customers' reaction on them in near real time (Fabijan et al., 2016). Big players [e.g., Amazon (Dmitriev et al., 2016), Facebook (Bakshy et al., 2014), Google (Hohnhold et al., 2015), Netflix (Amatriain and Basilico, 2012), or Uber (Deb et al., 2018)] as well as smaller companies have been using A/B testing as a scientifically grounded way to evaluate changes and comparing different alternatives (Deng et al., 2016). And, in the last years, the rapid rise of A/B testing has led to the emergence of multiple commercial testing platforms able to handle the implementation of these experiments (Dmitriev et al., 2017; Johari et al., 2017) that, according to the survey results presented in Fabijan et al. (2018b), are used by ~25% of web experimenters.

During the last decade, both scholars and practitioners have been publishing research articles, white papers and blog posts reporting recurrent pitfalls observed in their organizations (Crook et al., 2009; Kohavi et al., 2014; Dahl and Mumford, 2015; Dmitriev et al., 2017). In the specific case of evaluating web personalization and recommender systems, some of these pitfalls become especially recurrent, obscuring the interpretation of results or inducing invalid conclusions. Typically, most of the publications came from big digital companies, such as Microsoft (Dmitriev et al., 2017), Google (Hohnhold et al., 2015), Facebook (Bakshy et al., 2014), Uber (Deb et al., 2018), or Netflix (Amatriain and Basilico, 2012; Su and Yohai, 2019). However, both small-to-medium companies and also big traditional companies are now adopting website experimentation initiatives (Olsson et al., 2017; Fabijan et al., 2018a). From the observation of some of those initiatives in companies of the automotive sector (commonly seen as traditional industrial companies), we identified and reported some critical pitfalls for the reliability of AB tests that were repeated with worrying regularity (Esteller-Cucala et al., 2019).

The objective of this paper is to analyze if the pitfalls identified in the automotive industry are still present across industries. Specifically, we focus on pitfalls that are specially damaging or likely to appear when evaluating personalization features.

The list of pitfalls studied and presented in this paper was firstly obtained from the observation in a company of the automotive sector, and also, the commented pitfalls are limited to the ones considered basic for the implementation of a testing initiative (Kohavi et al., 2009).

In this paper we discuss a list of five experimentation pitfalls. In order to obtain the list, different information sources were used.

2. METHODS

In order to obtain the list of pitfalls on AB testing we suggest a mixed approach. To this effect, several procedures have been used.

The result of the three first data gathering methods were already shared in a previous work (Esteller-Cucala et al., 2019). In summary, these methods were:

1. General literature review on the topic of AB testing. From this review we obtained the first draft-list of pitfalls.
2. The active participation in a website testing project of a company in the automotive sector let us gather several data from their testing practices. The analyzed company works with multiple websites. At the time of study, more than 10 websites (managed by different teams) were being AB tested.

The data collection in this case consisted in test reports and participative observation.

3. In order to examine if the detected pitfalls are specific of the firstly observed company or generalizable across companies of the same sector, the observation was extended to other automotive companies. With this purpose, we collected data from other seven companies in the automotive sector. In this case, the data collection included summaries of their testing projects (in five of the companies), group meetings with six of the companies and open answer surveys in three of the companies.

After those three first steps, we had a list of pitfalls, identified the regular testing practices of real case companies, that was consistent with the literature. In order to study if these pitfalls should be a general concern across sectors, the observation was extended to companies in sectors other than automotive. In this case, the data collection included attending to open presentation of companies explaining their testing initiatives and a set 18 open-ended interviews.

The interviews consisted of, first, three demographic questions in order to know the sector of the company, the position of the respondent, the number of yearly AB tests run and the use of any commercial experimentation tool (the name of the participant as well as the name of the company were kept anonymous). Second, a set of seven questions were made in order to explore the standard experimentation routine in the respondent's company. The specific questions were oriented to inquire about each of the testing pitfalls of the list (without explicitly mentioning the pitfalls). In order to test if the questionnaire was correctly designed and the questions were correctly formulated to detect each of the pitfalls, a pilot respondent was surveyed. The pilot respondent was working in a company with several publications on the topic, with it, the expected answers were known beforehand.

3. RESULTS

As previously said, in this paper we are going to present and discuss a set of pitfalls that, even if they need to be kept in mind for any A/B test, they are more likely to appear when trying to assess the effect of personalized features. The pitfalls commented in this section are not only including statistical issues but also testing misconceptions or bad practices.

3.1. Evaluation Metrics Selection

According to different reports, marketers expect personalization effect in terms of visitors engagement, customer experience, brand perception and customer loyalty. However, they declare to be measuring the effects of personalization via improvements in conversion rates, click-through rates, revenues and page views, among others (Adobe, 2013; Benlian, 2015; Researchscape International and Evergage, Inc.). The importance of choosing an evaluation metric that really reflects business objectives is not a distinguishing concern of personalization feature experimenters, but one of the general key challenges for organizations that run controlled experiments (Kohavi et al., 2012). Experiments should be evaluated using metrics that reflect business objectives (Dahl and Mumford, 2015), and at the same time be understandable (as simple as possible to interpret the results), interesting to optimize, representative of good website performance (this is not giving positive results when the user experience is worsening) (Crook et al., 2009; Kohavi et al., 2014).

All in all, the evaluation metrics play a key role throughout the experimentation life cycle (design, running, overall evaluation and final decision) (Dmitriev et al., 2017). Therefore, it is recommended for experimenters to keep one single evaluation metric per experiment (Emily Robinson, 2018), agreed upfront (Kohavi et al., 2007), and kept during the whole test (Keser, 2018). Adding secondary objectives to monitor other relevant metrics or to compute complex predictors of long-term results can be a good practice as long as there is a clear unique and fix evaluation metric experiment.

In order to understand the evaluation metrics selection procedures of the different interviewees the question "*How do you choose the evaluation metrics of your tests?*" was directly asked. From both the observation and the interviews results, we can see how, almost every company is using more than one metric for the evaluation of their tests (except from the respondents working for experimentation consulting firms). Most of the respondents report that their companies combine general objectives of the organization and specific goals depending on the test details. With it, declaring a winner version of the test or deciding if the hypothesis is validated can become a difficult task and the final conclusions of the test might be left to the personal interpretation, which is the opposite of what a web testing initiative should stand for Kohavi et al. (2007). Moreover, the results are in line with the Experimentation Growth Model (Fabijan et al., 2018a). The companies with greater experience on AB testing report the use of stable metrics along their experiments, while companies with less experience report sets of evaluation metrics highly dependent on the specific experiment.

The evaluation metric selection might not seem a testing pitfall itself, however, we consider it the cornerstone of an online controlled experiment. If the unique evaluation metric of the experiment is not selected properly, both the utility and the validity of the test can be doubtful.

3.2. Determination of the Experiment Length

When using frequentist statistical approaches, the specific length (in time) of the experiment can not be determined in advance, it can only be estimated given a minimum experiment sample size and a predicted average of users (or any other test unit) per time unit. To determine the sample size of the test upfront is one of the most basic premises given for online controlled experiments. However, we have seen how numerous teams continuously monitor their experiments and stop them before the sample size is reached. Accordingly, this is one of the first advices that testing experts give in their papers and online blogs (Kohavi et al., 2007, 2014; Dahl and Mumford, 2015; Dmitriev et al., 2017; Emily Robinson, 2018). Reaching a specific minimum sample size before being able to obtain any result is one of the requirements of the Null Hypothesis Statistical Testing (NHST); nonetheless, this pitfall could turn irrelevant by changing to another statistical interpretation of the results, such as using Bayesian Hypothesis Testing or Sequential Hypothesis Testing (Deng et al., 2016; Johari et al., 2017; Su and Yohai, 2019) which have been attracting research interest as alternatives to NHST and are already used by some commercial testing tools [e.g., VWO and AB Tasty are based on Bayesian calculations (Stucchio, 2015; Wassner and Brebion, 2018) and Optimizely uses Sequential hypothesis testing (Rusonis and Ren, 2018)]. However, both the performed observations, interviews and previous authors report that frequentist approaches are still the most commonly used for A/B testing (e.g., Kohavi et al., 2007; Emily Robinson, 2018).

It is important to note that this pitfall is not only related to the early stopping of the experiment but also with the post-test segmentation. In order for the results to be valid, the minimum sample size required for the analysis is calculated. This is, if the minimum calculated sample size is X, the sample size of any post-test segmentation which is smaller than X is not be valid (Keser, 2018).

On the other side using much longer samples than needed can also arise in some experiment difficulties (Dmitriev et al., 2016). Sometimes, constrains to the experiment length are set in order to dissipate temporal effects, such as hour-of-day effects (Su and Yohai, 2019), day-of-week effects (Kohavi et al., 2007), business cycles (QuickSprout, 2019), or seasonality effects (Dmitriev et al., 2016). However, these effects might not impact all the organization or all the tests (Su and Yohai, 2019). In the specific case of a personalized feature depending on temporal factors (e.g., hour of the day) experimenters should consider whether to make a generalization or a case dependent experiment.

For the observed cases of the automotive sector, this was a relevant pitfall because there is only one of the observed companies using a non-frequentist approach. However, no-companies where calculating the sample size beforehand. Regarding the interviewees from companies form other sectors, there is a mix of companies calculating and not calculating the sample size required for the test in advance.

3.3. Multiple Comparison Problem

Even if the simplest case of A/B testing is considered when comparing only two variants (one against the control), there is no limit of variants to be compared in a single A/B test (also known as A/B/n test). For example, a common case in personalization is to test complex differences between variants, for this, one recommended approach is to test a collection of different variants including small or independent changes in order to be more precise in determining the specific effects of each variation included (Kohavi et al., 2014). This might also be the case when trying to adjust the personalization algorithms' parameters (Letham et al., 2018) or the individualization degree of the personalization (Arora et al., 2008). In this cases, testing a set of different variants is a good practice, even thought there is a statistical consideration to keep in mind when including more than two variants in an A/B test. When the sample size is calculated for a given significance level (e.g., 10%, equivalent to a 90% confidence level) each comparison has a false positive rate equal to the significance level. If we make multiple comparisons within the same test, the whole-test false positive rate is higher. For example, when trying to compare among 15 variants, the chance of getting a false positive (51%) is almost equivalent to flipping a coin and getting a head (Esteller-Cucala et al., 2019). Moreover, this effect should be taken into account any time that there is more than one comparison in the test (e.g., if more than one metrics monitored within the test or if the test is studied separately for different user segments). Nevertheless, some adjustments have been proposed in the literature (e.g., Bonferroni correction) in order to avoid this pitfall (Kohavi et al., 2007; Dahl and Mumford, 2015; Emily Robinson, 2018).

In order to see if the multiple comparison problem was an experimentation pitfall generally affecting to companies both in the automotive sector and in other sectors, the interview directly included a question asking if experiments with more than two variant were performed within the interviewee's company and, in if this was the case, respondents were asked if any criteria was used in order to adapt the experiment length. The results show that even if not all companies are familiar with more-than-two-variants experiments (specially the observed companies of the automotive sector), it is an extended practice and two thirds of the participants are testing with more than two variants. However, only a minority of respondents were aware of any existent corrections to be applied when conducting multiple comparison tests (apart from the consultancy companies).

3.4. Balance Among Experiment Samples

As above mentioned, the main objective with an A/B test is to establish a causal relationship between the test condition and a measurable change in some evaluation metric. This causal relationship is based on the premise that any external alteration to the metrics (except the tested ones) are controlled by the randomization and balanced between the test variants (Zhao and Zhao, 2016). Even if this balance condition is necessary for the test to be reliable, there is still lots of practitioners dismissing its importance. The unbalanced sampling refers to the situation where the split of users between variants does not satisfy the expected ratio (Dmitriev et al., 2017; Emily Robinson, 2018). In extremely unbalanced tests problems, such as the Simpson's paradox might appear (Crook et al., 2009). Due to this, unbalanced sampling is one of the most commented pitfalls (Crook et al., 2009; Dmitriev et al., 2017; Emily Robinson, 2018).

Some unbalance common causes are, for example, changing the sample ratio during the experiment (e.g., using ramp-ups to activate the test), post-test segmentation, post-test grouping of samples tested with different ratios or bugs in the implementation (e.g., a bug that affects only to a specific browser) (Kohavi et al., 2007; Crook et al., 2009; Keser, 2018).

Even though unbalanced sampling is a common pitfall in A/B testing, in websites where personalization is used it gets even more common (both when testing personalized or non-personalized features) (Das and Ranganath, 2013). For the specific case of personalization using *monitoring segments* is recommended. This is, to use segments that are not going to be used for making decisions about the result itself but to ensure that all the relevant distinguishable groups included in the personalization algorithms are distributed between variants according to the test ratio (e.g., segments based in scoring intervals). This technique is known as the stratified sampling (Urban et al., 2016; Keser, 2018).

As seen, there are different causes for unbalanced sampling. In order to not induce specific answers from the interviewees the questions regarding this pitfall were focused on two of the possible causes. First, we asked to the participants if they were using ramp-ups or other secure actuation methods in their tests. The result show that almost no respondents are using these methods, so we can conclude that they are not unbalancing the sampling this way. Second, we asked to the participants if they were regularly using AA tests in order to validate their testing tools [recommended practice to detect bugs that cause unbalance (Zhao and Zhao, 2016)]. The result show that <50% of the respondents report using these kind of validation tests in a regular basis. Even if we know that this pitfall appears in the automotive sector and is consistent with the literature, with the previous two questions we can not extract a conclusion about the generalizability of this pitfall.

3.5. Blind Adoption of Good Results

Even if A/B testing is one of the simplest evaluation techniques used for the evaluation of website performance (Knijnenburg, 2012; Bakshy et al., 2014), there are many variables that can affect the results. When the result of a test is unexpectedly bad (e.g., the new feature being tested under-performs by long the previous one) a frequent response is to look for the bug. On the contrary, this behavior is not as common when the unsuspected result is good. In the literature, this is known as "failing to apply Twyman's Law" (Dmitriev et al., 2017). A similar case is when a borderline p-value is given as a result from a test (Kohavi et al., 2014). But also, there is a common practice of activating new variants after a non-significant test result because "it doesn't hurt" (Emily Robinson, 2018). Even if each organizations might have different results, authors claim that only one-third of the experiments performed in their company improved the metrics they were designed to improve (Kohavi et al., 2014).

When thinking on the scientific rigor assumed for web experimentation, one may presume that this specific pitfall might be unlikely to happen in real organizations. However, as reported in Esteller-Cucala et al. (2019), we had the chance to see several times how some tests are prepared with high expectations of obtaining a specific result. After the collection of interviews we have seen that approximately half of the interviewees companies directly apply the new variant in case of a winning result. Seen the number of pitfalls that are not commonly considered when A/B testing, directly applying the results without more test iterations might result in the activation of false winners (false positive results), borderline p-values, insufficient sample sizes tested and so on. Some other respondents report an analysis of secondary metrics to decide whether to activate the winner variant or not.

In the specific case of testing personalization, it might be more difficult to deduct whether a given result makes sense or not, making it easier for some incorrect results to go unnoticed. For these reasons, even if the blind adoption is general a pitfall in A/B testing, it is even more likely to appear in the specific case of testing personalized features. Considering the double-check (or even *double-test*) of the test results might, in some cases, not only be a good practice but also a requirement especially in tests reporting unexpected or borderline results.

4. DISCUSSION AND CONCLUSIONS

While there is value in evaluating personalized features by means of online controlled experiments (A/B tests), there are some pitfalls to bear in mind while testing websites. In this paper, we discuss some critical AB testing pitfalls that were firstly identified in automotive companies and may compromise the validity of their experiments. Moreover, the analysis is then extended to study the presence of this testing pitfalls in companies from sectors other than automotive (always keeping the focus of the study on small-to-medium companies and also big traditional companies with relatively recent adoption of web testing initiatives). As a result, we presented five pitfall topics and commented their presence in the different sectors studied.

After a decade of publications from expert practitioners and big digital companies, the most basic and critical pitfalls are substantially well-documented. Despite this, companies adopting AB testing seem not being completely aware of this testing pitfalls. As seen in the results of this study, most of the respondent companies have not a clear procedure for the selection of their evaluation metrics, which is the starting point of an AB test. Moreover, a remarkable number of the surveyed companies directly apply winner results without further analysis of the test (blind adoption of good results) and are not aware of the multiple comparison problem and its possible corrections to take it into account. However, even if there is still a noteworthy proportion of companies not determining the experiment length beforehand (when using frequentist statistical approaches) the results for the general industries surveys are better than in the companies of the automotive sector. Finally, regarding the pitfalls with the balance among experiment samples, the answers gotten from the survey are not clear enough to extract a conclusion. With it, our results show that there are some basic AB testing pitfalls, well-known by scholars and big digital companies, that are present in the experimentation initiatives of companies relatively inexpert with AB testing.

As previously stated, the list of pitfalls included in this study is by no means the complete list of possible pitfalls that may appear when performing AB tests or even the complete list of pitfalls that can be collected by reviewing the literature. Other pitfalls are still commonly seen in companies and may appear while running specific tests. Some examples are: not considering temporal effects on the user behavior (e.g., holiday seasons or Valentine's Day), neglecting novelty effects or ignoring temporal cycles (both business or calendar cycles) (Kohavi et al., 2007; Dmitriev et al., 2016, 2017; Weinstein, 2019). Even though those pitfalls are also important and should be studied in order to verify the reliability of each result, they might not apply for each test and company (Su and Yohai, 2019). The pitfalls commented in this paper are limited to the ones observed in a specific company of the automotive sector (and then validated with other companies) and the ones considered critical for the validity of the test. If they are not understood and addressed properly, these pitfalls might invalidate not only specific test but the entire testing initiative of a company.

However, this study has some limitations, here we point out four of them. First, the list of testing pitfalls commented in this paper was firstly focused on the automotive industry, therefore, some important pitfalls for other sectors might be missing. Second, the list here presented is not a complete list of possible web AB testing pitfalls, but a list of the observed ones that are considered the most basic and critical for the global testing initiative of a company. Third, the total number of analyzed companies is not large enough to statistically determine the generalizability of each of the presented testing pitfalls. Further research could extend the study to a larger group of companies. Finally, this work is only focused on examining the presence of these testing pitfalls across the industry. However, the reasons why these testing pitfalls can still be found inside the companies, despite the large body of knowledge available on how to identify and avoid them, are not studied and could be addressed on further research.

Additionally, further work needs to be done in the experimentation procedures organizations use to evaluate their personalization efforts. With it, we propose to organizations to construct their own evaluation framework. This is, inspired in the most common pitfalls reported in A/B testing, organization could set the conditions for their teams to experiment. This framework should include for example, the criteria for the evaluation metrics selection, the criteria to be used in order to determine the experiments length (not determining the specific length but setting the criteria to determine it), post-test segmentation criteria and results adoption criteria.

With this work, we aim to increase the experimenters' awareness on those pitfalls. And also, to attract the attention of persuasive technology scholars on the gap between academia advances on the personalization field and its adoption on the industry.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. The interview data might contain personal data that has been omitted in the article.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ME-C, VF, and DV contributed to the design. ME-C implemented the research, the analysis of the results and the writing of the first manuscript. VF and DV supervised the project. ME-C, VF, and DV contributed to the final version of the manuscript.

FUNDING

This work was partially supported in part under the Industrial Doctorate Grand DI 052/2016 (Secretaria d'Universitats i Recerca, Generalitat de Catalunya).

ACKNOWLEDGMENTS

This work was partially based in a paper previously disseminated in *The 27th ACM Conference On User Modeling, Adaptation and Personalization* workshop and adjunct proceedings ACM ISBN 978-1-4503-6711-0 (Esteller-Cucala et al., 2019).

REFERENCES

Adobe (2013). *Click Here: The State of Online Advertising*. Technical report, Adobe Systems Incorporated.

Amatriain, X., and Basilico, J. (2012). *Netflix Recommendations: Beyond the 5 Stars (Part 2)*. Available online at: https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5 (accessed April 2, 2020).

Arora, N., Dreze, X., Ghose, A., Hess, J. D., Iyengar, R., Jing, B., et al. (2008). Putting one-to-one marketing to work: personalization, customization, and choice. *Market. Lett*. 19, 305–321. doi: 10.1007/s11002-008-9056-z

Bakshy, E., Eckles, D., and Bernstein, M. S. (2014). "Designing and deploying online field experiments," in *International World Wide Web Conference Committe (IW3C2)* (Seoul).

Benlian, A. (2015). Web personalization cues and their differential effects on user assessments of website value. *J. Manag. Inform. Syst*. 32, 225–260. doi: 10.1080/07421222.2015.1029394

Bleier, A., De Keyser, A., and Verleye, K. (2017). "Customer engagement through personalization and customization," in *Customer Engagement Marketing, Chapter 4*, eds R. W. Palmatier, V. Kumar, and C. M. Harmeling (Cham: Palgrave Macmillan), 75–94. doi: 10.1007/978-3-319-61985-9_4

Choi, B., Kwon, O., and Shin, B. (2017). Location-based system: comparative effects of personalization vs ease of use. *Telem. Inform*. 34, 91–102. doi: 10.1016/j.tele.2016.04.011

Coelho, P. S., and Henseler, J. (2012). Creating customer loyalty through service customization. *Eur. J. Market*. 46, 331–356. doi: 10.1108/03090561211202503

Crook, T., Frasca, B., Kohavi, R., and Longbotham, R. (2009). "Seven pitfalls to avoid when running controlled experiments on the web," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD '09* (Paris), 1105. doi: 10.1145/1557019.1557139

Dahl, J., and Mumford, D. (2015). *Nine Common A / B Testing Pitfalls and How to Avoid Them*. Technical report, Adobe Systems Europe Ltd.

Das, A., and Ranganath, H. (2013). "When web personalization misleads bucket testing," in *UEO '13: Proceedings of the 1st Workshop on User Engagement Optimization* (San Francisco, CA), 17–20. doi: 10.1145/2512875.2512879

Deb, A., Bhattacharya, S., Gu, J., Zhou, T., Feng, E., and Liu, M. (2018). *Under the Hood of Uber's Experimentation Platform*.

Demangeot, C., and Broderick, A. J. (2016). Engaging customers during a website visit: a model of website customer engagement. *Int. J. Retail Distrib. Manag*. 44, 814–839. doi: 10.1108/IJRDM-08-2015-0124

Deng, A., Lu, J., and Chen, S. (2016). "Continuous monitoring of A/B tests without pain: optional stopping in Bayesian testing," in *Proceedings-3rd IEEE*

International Conference on Data Science and Advanced Analytics, DSAA 2016 (Montreal, QC), 243–252. doi: 10.1109/DSAA.2016.33

Dmitriev, P., Frasca, B., Gupta, S., Kohavi, R., and Vaz, G. (2016). "Pitfalls of long-term online controlled experiments," in *2016 IEEE International Conference on Big Data (Big Data)* (Washington, DC), 1367–1376. doi: 10.1109/BigData.2016.7840744

Dmitriev, P., Gupta, S., Dong Woo, K., and Vaz, G. (2017). "A dirty dozen: twelve common metric interpretation pitfalls in online controlled experiments," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD '17* (Halifax, NS), 1427–1436. doi: 10.1145/3097983.3098024

Esteller-Cucala, M., Fernandez, V., and Villuendas, D. (2019). "Experimentation pitfalls to avoid in A/B testing for online personalization," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, UMAP'19 Adjunct* (New York, NY: ACM), 153–159. doi: 10.1145/3314183.3323853

Fabijan, A., Dmitriev, P., McFarland, C., Vermeer, L., Holmström Olsson, H., and Bosch, J. (2018a). Experimentation growth: evolving trustworthy A/B testing capabilities in online software companies. *J. Softw. Evol. Process* 30:e2113. doi: 10.1002/smr.2113

Fabijan, A., Dmitriev, P., Olsson, H. H., and Bosch, J. (2018b). "Online controlled experimentation at scale: an empirical survey on the current state of A/B testing," in *44th Euromicron Conference on Software Engineering and Advanced Applications* (Prague), Vol. 7, 57–61. doi: 10.1109/SEAA.2018.00021

Fabijan, A., Olsson, H. H., and Bosch, J. (2016). "The lack of sharing of customer data in large software organizations: challenges and implications," in *17th International Conference on Agile Processes in Software Engineering and Extreme Programming, XP 2016* (Edinburgh, UK), Vol. 1, 334. doi: 10.1007/978-3-319-33515-5_4

Govind, N. (2017). *A/B Testing and Beyond: Improving the Netflix Streaming Experience With Experimentation and Data Science*. Available online at: https://netflixtechblog.com/a-b-testing-and-beyond-improving-the-netflix-streaming-experience-with-experimentation-and-data-5b0ae9295bdf (accessed April 2, 2020).

Hohnhold, H., O'Brien, D., and Tang, D. (2015). "Focusing on the long-term: it's good for users and business," in *Proceedings 21st Conference on Knowledge Discovery and Data Mining*, Vol. 2015 (Sydney, NSW: ACM). doi: 10.1145/2783258.2788583

Johari, R., Koomen, P., Pekelis, L., and Walsh, D. (2017). "Peeking at A/B tests: why it matters, and what to do about it," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD '17* (Halifax, NS), 1517–1525. doi: 10.1145/3097983.3097992

Kaptein, M., Markopoulos, P., De Ruyter, B., and Aarts, E. (2015). Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. *Int. J. Hum. Comput. Stud.* 77, 38–51. doi: 10.1016/j.ijhcs.2015.01.004

Kaptein, M., and Parvinen, P. (2015). Advancing e-commerce personalization: process framework and case study. *Int. J. Electron. Commerce* 19, 7–33. doi: 10.1080/10864415.2015.1000216

Keser, C. V. (2018). *The Top 3 Mistakes That Make Your A/B Test Results Invalid*. Available online at: https://www.widerfunnel.com/3-mistakes-invalidate-ab-test-results/ (accessed April 2, 2020).

Knijnenburg, B. P. (2012). "Conducting user experiments in recommender systems," in *Proceedings of the Sixth ACM Conference on Recommender Systems-RecSys '12* (Dublin, UK), 3. doi: 10.1145/2365952.2365956

Kohavi, R., Deng, A., Frasca, B., Longbotham, R., Walker, T., and Xu, Y. (2012). "Trustworthy online controlled experiments: five puzzling outcomes explained," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD '12* (Beijing), 786. doi: 10.1145/2339530.2339653

Kohavi, R., Deng, A., Longbotham, R., and Xu, Y. (2014). "Seven rules of thumb for web site experimenters," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD '14* (New York, NY), 1857–1866. doi: 10.1145/2623330.2623341

Kohavi, R. M., Henne, R. M. M., and Sommerfield, D. M. (2007). "Practical guide to controlled experiments on the web: listen to your customers not to the HiPPO," in *KDD-2007: 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Jose, CA), 959–967. doi: 10.1145/1281192.1281295

Kohavi, R. M., Longbotham, R., Sommerfield, D. M., and Henne, R. M. M. (2009). Controlled experiments on the web : survey and practical guide. *Data Mining Knowl. Discov*. 18, 140–181. doi: 10.1007/s10618-008-0114-1

Kwon, K., Cho, J., and Park, Y. (2010). How to best characterize the personalization construct for e-services. *Expert Syst. Appl.* 37, 2232–2240. doi: 10.1016/j.eswa.2009.07.050

Lee, G., and Lin, H. (2005). Customer perceptions of e-service quality in online shopping. *Int. J. Retail Distrib. Manag.* 33, 161–176. doi: 10.1108/09590550510581485

Letham, B., Karrer, B., Ottoni, G., and Bakshy, E. (2018). Constrained Bayesian optimization with noisy experiments. *Bayesian Anal.* 14, 495–519. doi: 10.1214/18-BA1110

Mahmoud, M. A., Kastner, A. N. A., and Lartey, J. O. (2017). Internet-based relationship marketing: a sub-saharan african perspective. *J. Relat. Market*. 16, 179–196. doi: 10.1080/15332667.2017.1349555

Mendoza, J. M. G., and Marasinghe, A. (2013). "Kansei color concepts in interface design for Mexican and Japanese e-commerce websites," in *Proceedings-2013 International Conference on Biometrics and Kansei Engineering, ICBAKE 2013* (Tokyo), 154–159. doi: 10.1109/ICBAKE.2013.62

Oinas-Kukkonen, H., and Harjumaa, M. (2008a). "A systematic framework for designing and evaluating persuasive systems," in *Persuasive Technology, 3rd International Conference, PERSUASIVE 2008* (Oulu). doi: 10.1007/978-3-540-68504-3_15

Oinas-Kukkonen, H., and Harjumaa, M. (2008b). "Towards deeper understanding of persuasion in software and information systems," in *First International Conference on Advances in Computer-Human Interaction* (Sainte Luce). doi: 10.1109/ACHI.2008.31

Olsson, H. H., Bosch, J., and Fabijan, A. (2017). Experimentation that matters: a multi-case study on the challenges with A/B testing. *8th Int. Conf. Softw. Bus.* 304, 179–185. doi: 10.1007/978-3-319-69191-6_12

Piccoli, G., Lui, T.-W., and Grün, B. (2017). The impact of IT-enabled customer service systems on service personalization, customer service perceptions, and hotel performance. *Tourism Manag.* 59, 349–362. doi: 10.1016/j.tourman.2016.08.015

QuickSprout (2019). *15 A/B Testing Mistakes to Avoid*. Available online at: https://www.quicksprout.com/ab-testing-mistakes/ (accessed April 2, 2020).

Rashid, R. A., Nadiah, S., Sapari, A., Majid, N. H., Wahab, Z., Yunus, K., et al. (2016). Analysis of persuasive elements in multi-level marketing official websites. *Int. J. Appl. Bus. Econ. Res.* 14, 1139–1146.

Researchscape International and Evergage, Inc. (2018). *2018 Trends in Personalization*. Technical report. Researchscape International; Evergage, Inc.

Robinson, E. (2018). *Guidelines for A/B Testing*. Available online at: https://hookedondata.org/guidelines-for-ab-testing/ (accessed April 2, 2020).

Rusonis, S., and Ren, J. (2018). *A Practical Guide to Statistics for Online Experiments*.

Salonen, V., and Karjaluoto, H. (2016). Web personalization: the state of the art and future avenues for research and practice. *Telem. Inform.* 33, 1088–1104. doi: 10.1016/j.tele.2016.03.004

Spiliopoulou, M. (2000). Web usage mining for web site evaluation. *Commun. ACM* 43, 127–134. doi: 10.1145/345124.345167

Stucchio, C. (2015). *Bayesian A/B Testing at VWO*.

Su, G., and Yohai, I. (2019). *Improving Experimentation Efficiency at Netflix with Meta Analysis and Optimal Stopping*. Available online at: https://netflixtechblog.com/improving-experimentation-efficiency-at-netflix-with-meta-analysis-and-optimal-stopping-d8ec290ae5be (accessed April 2, 2020).

Tam, K. Y., and Ho, S. Y. (2005). Web personalization as a persuasion strategy: an elaboration likelihood model perspective. *Inform. Syst. Res.* 16, 271–291. doi: 10.1287/isre.1050.0058

Urban, S., Sreenivasan, R., and Kannan, V. (2016). *It's All A/Bout Testing: The Netflix Experimentation Platform*. Available online at: https://netflixtechblog.com/its-all-a-bout-testing-the-netflix-experimentation-platform-4e1ca458c15 (accessed April 2, 2020).

Wassner, H., and Brebion, A. (2018). *Demystifying A/B Testing Statistics*.

Weinstein, A. (2019). *The Dark Side of A/B Testing*. Available online at: https://venturebeat.com/2019/04/13/the-dark-side-of-a-b-testing/ (accessed April 2, 2020).

Xu, J. D., Benbasat, I., and Cenfetelli, R. T. (2014). The influences of online service technologies and task complexity on efficiency and personalization. *Inform. Syst. Res*. 25, 420–436. doi: 10.1287/isre.2013.0503

Yen, B., Hu, P. J.-H., and Wang, M. (2007). Toward an analytical approach for effective Web site design: a framework for modeling, evaluation and enhancement. *Electron. Commerce Res. Appl*. 6, 159–170. doi: 10.1016/j.elerap.2006.11.004

Zhao, Y., and Zhao, D. (2016). The personalization willingness paradox: an empirical evaluation of sharing information and prospective benefit of online consumers. *Rev. Iber. Sist. Tecnol. Inform*. 2016, 30–44.

Conflict of Interest: ME-C and DV were employed by the company SEAT, S.A.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

SYSTEMATIC REVIEW
published: 28 April 2020
doi: 10.3389/frai.2020.00007

Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review

Noora Aldenaini [1,2], Felwah Alqahtani [1,3], Rita Orji [1] and Srinivas Sampalli [1]*

[1] *Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada,* [2] *Department of Computer Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia,* [3] *Department of Computer Science, King Khalid University, Abha, Saudi Arabia*

Persuasive technology (PT) is increasingly being used in the health and wellness domain to motivate and assist users with different lifestyles and behavioral health issues to change their attitudes and/or behaviors. There is growing evidence that PT can be effective at promoting behaviors in many health and wellness domains, including promoting physical activity (PA), healthy eating, and reducing sedentary behavior (SB). SB has been shown to pose a risk to overall health. Thus, reducing SB and increasing PA have been the focus of much PT work. This paper aims to provide a systematic review of PTs for promoting PA and reducing SB. Specifically, we answer some fundamental questions regarding its design and effectiveness based on an empirical review of the literature on PTs for promoting PA and discouraging SB, from 2003 to 2019 (170 papers). There are three main objectives: (1) to evaluate the effectiveness of PT in promoting PA and reducing SB; (2) to summarize and highlight trends in the outcomes such as system design, research methods, persuasive strategies employed and their implementaions, behavioral theories, and employed technological platforms; (3) to reveal the pitfalls and gaps in the present literature that can be leveraged and used to inform future research on designing PT for PA and SB.

Keywords: persuasive technology, persuasive strategies, behavior theory, targeted audience, targeted outcomes, physical activity, sedentary behavior, health

OPEN ACCESS

Edited by:
Styliani Kleanthous, Open University of Cyprus, Cyprus

Reviewed by:
Olga C. Santos, National University of Distance Education, Spain
Cesar Collazos, University of Cauca, Colombia

****Correspondence:***
Noora Aldenaini
nr412864@dal.ca

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *17 October 2019*
Accepted: *21 February 2020*
Published: *28 April 2020*

Citation:
Aldenaini N, Alqahtani F, Orji R and Sampalli S (2020) Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. Front. Artif. Intell. 3:7. doi: 10.3389/frai.2020.00007

1. INTRODUCTION

In recent years, our way of life has become increasingly sedentary, which is a significant public health issue. Sedentary behavior (SB) is defined as any awake behavior that has an energy expenditure $\leq$1.5 metabolic equivalent (METs). This may include a sitting, lying, or reclining posture such as watching television and working at a desk [1]. When we compare our life to previous generations, it is clear that our life has become more sedentary. For example, some individuals are spending more time in environments that limit physical activity (PA) and require prolonged sitting. A sedentary lifestyle is associated with health complications such as obesity, diabetes, cancer, and cardiovascular diseases, among other conditions [2]. Thus, reducing SB and increasing PA has been the focus of much PT. There is a need to understand how persuasive technology (PT) has been used to promote health and prevent disease by targeting certain behaviors in the individual that promote their PA and reduce SB.

Over the years, considerable research has designed and used PT to promote PA and discourage SB. Thus, it is important to understand and evaluate the effectiveness of these PT at achieving their intended outcome of reducing the health risks associated with a sedentary lifestyle by promoting PA.

Therefore, in this paper we aim to achieve three main objectives: (1) to evaluate the effectiveness of PTs used to promote PA and reduce SB; (2) to summarize and highlight trends in the outcomes such as system design, research methods, persuasive strategies employed and their implementations, behavioral theories, and employed technological platforms; and (3) to reveal pitfalls and gaps in the present literature that could be leveraged and used to inform the design of PTs targeting physical activity. To achieve this, we conducted a systematic review of 170 research papers to identify and evaluate the effectiveness of PT for promoting PA and discouraging SB using the Persuasive System Design (PSD) model [3] as shown in **Table 1**.

2. LITERATURE REVIEW

PT is a computer system that is designed to be interactive in a way that it can influence the attitude, beliefs, and behavior of the user to achieve a certain objective [4]. Fogg [5] further defined persuasive technology as "*the computing systems, devices, or applications intentionally designed to change a person's attitudes or behavior in a predetermined way*." The use of the term "persuasion" implies that the attitude and behavior of the user can be changed in a predetermined way in accordance with the plans and design intents of the persuasive technology's designer. Within the health domain, PTs can be used to either promote health and prevent disease, or to manage diseases and health conditions [6]. Many researchers have designed PT to help people to change their lifestyle and become more active. We present an overview of the literature review of PT interventions targeting both the SB and PA health domain.

2.1. Sedentary Behavior

There are many studies that have examined and evaluated the effectiveness of digital interventions in the health domain that aim to reduce SB for individuals.

The majority of the PT studies in the SB domain have targeted office workers and workplace interventions. For example, Wang et al. [7] conducted a systematic review to evaluate the use and effectiveness of PTs targeting SB in the work environment using the PSD model. They found that reminders were the most employed strategy to reduce SB. However, reminders alone have no substantial impact on SB reduction.

Similarly, Gardner et al. [8] reviewed 26 studies and identified the behavior change strategies employed in the SB interventions using behavior change techniques (BCTs). They examined the effectiveness of the identified strategies. Their findings revealed that problem-solving, self-monitoring, and reorganization of the social or physical environment were effective strategies in decreasing SB among adults.

There are other workplace interventions that are aimed at reducing SB. For example, Healy et al. [9] presented a review of 11 studies that aimed to reduce SB and offer a healthy work environment. They reinforced the implementation of motivational strategies (e.g., the use of a combination of several strategies, the increase in the number of breaks taken from sitting time, the focus on comfortable changes to people's workplace, the change to a healthy posture periodically, etc.) to decrease prolonged workplace sitting and mitigate the risks of such unhealthy behaviors. These strategies played an essential role in improving the individual health status in the workplace environment, increasing productivity, and decreasing absenteeism and injury costs.

Similarly, Shrestha et al. [10] reviewed a total of eight studies that aimed to reduce SB in the workplace. A total of 1,125 users who participated in the study were divided into intervention groups: policy changes, physical workplace changes, and information and counseling. The findings indicated that sit-stand desks were able to decrease sitting time at work, while the consequences of the information and counseling as well as policy changes were unpredictable. All eight selected review studies provided low-quality evidence due to the high risk of bias, poor research design, and small sample sizes.

Moreover, Chu et al. [11] showed evidence in their review paper for intervention effectiveness in decreasing SB in the workplace environment, especially for multi-component interventions (e.g., the installation of sit-stand workstations with the use of wearable activity trackers in combination with behavioral change strategies), and environmental strategies (e.g., the use of sit-stand workstations, treadmill desks, stationary cycle ergometers, and portable elliptical/pedal machines). They showed that the use of multi-component interventions was more promising than implementing educational/behavioral strategies alone. However, they did not compare the effectiveness of different behavior change techniques "strategies," as it is crucial to provide instructions and recommendations for PT design.

Addtionally, there are some studies that have evaluated the effectiveness of mobile applications in mitigating SB. For example, Dunn et al. [12] conducted a systematic review of persuasive strategies in 50 mobile applications (36 free apps, 14 paid apps) for reducing SB (e.g., sitting, laying on a bed, etc.) to identify the persuasive strategies employed in them using a taxonomy of 93 BCTs. The results showed that SB apps employed fewer persuasive strategies compared to PA mobile apps and other technology interventions in the health domains.

2.2. Physical Activity

Considerable studies have been focused in the area of analyzing the efficacy of PTs for promoting PA. Most of the PA interventions were mainly focused on using mobile applications and wearable devices technologies. McCallum et al. [13] examined 111 studies to evaluate PA promoting smartphone apps and wearable devices from different aspects: effectiveness, acceptability, engagement, and the implementation of rapid research designs. The results suggest the need to provide guidance to health and human-computer interaction (HCI) researchers in using more in-device sensors, user-logs, and rapid research designs.

Rao [14] provided a review paper on the usage of wearable activity monitoring devices for tracking and measuring PA in

TABLE 1 | Principles "strategies" of Persuasive System Design (PSD) Model [3].

Descriptions of PSD model strategies "Principles"	
Primary task support	
Reduction	The system has to decrease effort and strain that users consume when doing their target behavior. The reduction principle can be achieved by reducing a complex behavior into simple and easy tasks for users.
Tunneling	The system has to guide users in the attitude change process or experience by providing opportunities for action performance that makes user nearer to the target behavior.
Tailoring	The system has to offer tailored information for its user group according to their interests, needs, personality, or other factors related to the user group.
Personalization	The system has to provide personalized content and customized services for users.
Self-monitoring	The system has to give means for users to track and monitor their performance, progress, or status in accomplishing their goals.
Simulation	The system needs to give means for observing and noticing the connection between the cause and effect of users' behavior.
Rehearsal	The system must deliver means for rehearsing a target behavior.
Dialogue support	
Praise	The system has to deliver praise through images, symbols, words, videos, or sounds as an approach to give user feedback information regarding his/her behavior.
Rewards	The system should offer virtual rewards for users to provide credit for doing the target behavior. The virtual rewards come in different forms such as collecting points or trophies, and changing media elements (e.g., background, sounds, or avatar), etc.
Reminders	The system has to remind users to perform their target behavior while using the system.
Suggestion	The system has to suggest ways that users can achieve the target behavior and maintain performing behavior during the use of the system.
Similarity	The system must imitate its users in some particular manner, so the system should remind the users of themselves in a meaningful way.
Liking	The system should be visually attractive and contain a look and feel that meets its users' desires and appealing.
Social role	The system has to adopt a social role by supporting the communication between users and the system's specialists.
System credibility support	
Trustworthiness	The system has to give truthful, fair, reasonable, and unbiased information.
Expertise	The system has to offer information displaying experience, knowledge, and competence.
Surface credibility	The system must have a competent look and feel that portrays system credibility based on an initial assessment.
Real-world feel	The system must give information of the organization and/or the real individuals behind its content and services.
Authority	The system should refer to people in the role of authority.
Third-party endorsements	The system should deliver endorsements from well-known and respected sources.
Verifiability	The system has to give means to investigate the accuracy of the system content through external sources.
Social support	
Social learning	The system has to give a user the ability to observe other users and their performance outcomes while they are doing their target behavior.
Social comparison	The system should enable users to compare their performance with other users' performance.
Normative influence	The system has to have a feature for gathering together individuals that have identical objectives and let them feel norms.
Social facilitation	The system should enable a user to discern other users who are performing the target behavior along with him/her.
Cooperation	The system should offer the opportunity for a user to cooperate with other users to achieve the target behavior goal.
Competition	The system should allow a user to compete with other users. In the competition principle, there is a chance for winning or losing a race.
Recognition	The system has to offer public recognition (e.g., ranking) for users who do their target behavior.

older people. Rao suggested that wearable sensors are perfect for measuring PA intensity, step counts, and energy expenditure, however; there is a need to enhance the accuracy of measurement in this type of PA, non-ambulatory PA, and the spatial extent of PA.

There were other mobile applications and wearable tracker device-based interventions that targeted increasing PA. Stephens and Allen [15], in their systematic review, examined user satisfaction and the usefulness of smartphone applications and text messaging technology to support PA and weight loss. Seven articles published between 2005 and 2010 were included in their review paper. Their results indicated that all the technology interventions that included educational support or had more interventions showed greater effectiveness for smartphone and text messaging for weight loss and the increase of PA.

Similarly, a review of Lau et al. [16] assessed the success and quality of methods used in the information and communication technologies (ICTs)-based PA domain (e.g., Internet and mobile phones), specifically for children and adolescent populations. Nine studies (published between 2001 and 2009) were included and analyzed in their review article. These studies provided PA related to behavioral, psychosocial, and cognitive outcomes. Their findings showed the positive effects of ICTs in the PA domain for children and adolescents, especially when implemented with additional delivery methods (e.g., the face-to-face approach).

Tong and Laranjo [17] also wrote a review paper that characterized and assessed the effects of social features integration in mobile health (mHealth) interventions in promoting PA. They included 19 studies in their research, and their findings showed that social aspects were mostly employed to offer social support or comparison. Furthermore, some individuals were more motivated by social support and social competition, while others had concerns about social comparison. They found that social features may increase user engagement and increase users' PA levels; however, they also found it too difficult to determine the most effective features for increasing PA in mobile health technology due to the multi-component interventions of most of the studies they reviewed.

Hardeman et al. [18] conducted a systematic review of just-in-time adaptive interventions (JITAIs) in mobile health (mHealth) technology for PA to determine these interventions' effectiveness, feasibility, features, and acceptability. There were 19 papers included in the review, and 14 unique JITAIs were identified. Hardeman and colleagues emphasized that research into JITAIs' effectiveness in decreasing SB and increasing PA in its early stages, and there is a need for more evidence by endorsing the robust assessment of theory and evidence-based JITAIs.

Ehn et al. [19] provided a qualitative study of "elderly" users' experiences of using activity monitors to track and measure their performance for supporting PA in daily life. There were eight users involved in the qualitative study, and they perceived the wearable devices as easy to handle. Ehn and partners suggested that activity monitors can be used for motivating elderly people to adopt a good level of PA and to promote a healthy lifestyle. However, Ehn et al. identified areas that need development and enhancement such as usability, reliability, and content supporting successful BCTs to increase older people's engagements in PA.

Hamasaki [20] summarized studies (published between 2015 and 2018) to investigate the efficacy of using wearable devices, particularly mobile applications, to manage diabetes for diabetic patients. A total of four studies were included in the review paper. Hamasaki's review results showed that the use of accelerometers or pedometers increased PA by about 1 h weekly, while diabetes and obesity rates were not changed. He also found that smartphone applications are beneficial for encouraging PA and treating diabetes. Consequently, the use of wearable devices and smartphone apps by diabetic patients increases their interactions due to the self-monitoring, education, and coaching features implemented in these technologies. However, the author mentioned that there is still a need to investigate the most useful wearable devices that can be used by diabetics patients to track their PA level, heart rate, blood glucose level, blood pressure, and energy balance accurately and comfortably.

Bort-Roig et al. [21] introduced a systematic review paper of smartphones app for PA with a total of 26 articles published between 2007 and 2013. They showed proof on smartphones and their ability to measure and influencing PA. Moreover, they recommended working on identifying and having well-designed studies to help in evaluating the accuracy of PA measurements along with employing long-term assessments.

Matthews et al. [22] provided a systematic review of 20 articles for health behavioral-change of mobile apps, especially those apps aimed at promoting PA. The authors employed the PSD model for evaluating the inbuilt persuasive strategies of mobile apps in their reviewed articles. Their findings showed that the most commonly employed persuasive strategies were primary task support, social support, and dialogue support, while the least frequently employed was credibility support.

Ghanvatkar et al. [23] offered a scoping review of 48 studies to address the use of a personalization strategy for PA interventions, to recognize the different types of personalization, and to identify the user models employed for delivering personalization. Their review covered only the studies that implemented a personalization strategy in the design of the PT for PA regardless of the use of other persuasive strategies. The authors provided some recommendations and feedback for the researchers and developers of PTs (e.g., fitness devices, mobile apps) in the use of personalization strategies to increase PA.

Other studies have evaluated different PT interventions in encouraging PA. For instance, Almutari and Orji [24] presented an empirical review of 19 years (54 studies) of literature on PT for influencing PA. The authors included 54 papers (published from 2000 to 2019) in their report to assess the effectiveness of implementing social support strategies in PT for PA. They only included papers that focused mainly on employing the most frequently used social support strategies as social cooperation, social comparison, and social competition. Their findings suggest that PTs implementing socially-oriented strategies in the design of PT are considered successful tools to encourage and increase users' PA levels. The review papers conducted by Win et al. [25, 26] are other examples of a PT intervention in PA.

2.3. Studies Examining Both Physical Activity and Sedentary Behavior

This section includes the review papers that have focused on PT interventions in the area of both increasing PA and reducing SB.

A number of studies combined both PA and SB. For example, Prince et al. [27] provided a qualitative analysis of systematic review papers, including six studies in the PA and/or SB health domain. The authors aimed to provide a comparison of the efficacy of the interventions used on PA and/or SB to decrease the time spent sedentary in the adult population. Their findings indicate that a huge and clinically significant decrease in sedentary time can be achieved using interventions concentrating on reducing SB.

Schembre et al. [28] in their systematic review, evaluated data on the content features of feedback messaging employed in diet, PA, and SB interventions. The authors also created a practical framework to help developers to design just-in-time feedback for health behavior change in individuals. Approximately 31 studies were included in their review, in which 30 used personalized feedback, 24 employed goal-oriented feedback, and just 5 implemented actionable feedback. Furthermore, their results show that the feedback was often available, personalized, and actionable feedback with substantial behavior change outcomes.

Schoeppe et al. [29] investigated the effectiveness of health interventions that employ smartphone apps to enhance PA, SB, and diet in children and adult populations. Their systematic review examined twenty-seven studies published between 2006 and 2016. The results suggested that app-based interventions can be very useful in improving diet, PA, and SB. Furthermore, multi-component interventions seemed to be more promising and effective than stand-alone smartphone app interventions.

Yim and Graham [30] reviewed the literature on PA motivation and SB reduction by investigating the properties of digital exercise games. The authors introduced an exercise game called "Life is a Village" to demonstrate the exercise motivation needs and requirements for computer-aided exercise games.

The objective of the review paper by Lister et al. [31] was to identify and analyze the use of gamification in health and PA "fitness" apps in motivating users to adopt desirable and healthy behavior. Lister and partners examined health apps from the Apple App Store that were associated with diet and PA domains. The authors reviewed 132 apps and determined the top ten successful game elements, the top six essential health gamification elements, and the 13 most fundamental health behavior concepts. Their results indicated that the use of gamification in fitness and health apps was prevalent, and there was a lack of implementing behavior theory elements in the app industry.

It is obvious from the above literature review of the related work that some systematic studies have focused only on one specific domain, either PA or SB. Others have considered both fields of reducing SB and increasing PA while focusing on targeting a particular technology, population, or strategy. However, none of these studies have provided a comprehensive overview of the development and trends of PTs in PA and/or SB domains. For example, some reviews concentrated on a particular PT such as the use of smartphone apps, wearable devices, or games in promoting physical activity. Other papers focused only on reviewing studies that used one or a particular set of motivational strategies such as personalization or social support features, whereas another collection of papers focused on a specific target audience, such as children, elderly, or adults. Therefore, there is a need to provide a systematic review paper that offers a comprehensive overview of PTs in both PA and SB domain to bridge existing gaps from the review papers.

3. MATERIALS AND METHODS

The aim of this study is to evaluate the effectiveness of PT in reducing sedentary lifestyles and increasing the level of PA. The research questions of our systematic review paper are:

- To what extent are PTs effective in promoting PA and reducing SB?
- What are the outcomes' trends of employing PTs in promoting PA and reducing SB?
- What persuasive strategies were employed in designing PTs for PA and SB and how were they implemented?
- What are the pitfalls and gaps in the present literature on PT for PA and SB?
- What are the opportunities and recommendations for future PTs design?

We conducted a systematic review of 170 published papers in the PA and SB domains between 2003 and 2019. To achieve this, we used quantitative content analysis, a technique that enables the comparison, contrast, and categorization of data according to different themes and concepts, as adapted from Orji and Moffatt [4]. This entails collecting data in a rigorous way, paying special attention to the objectivity of the results. To retrieve articles for this review, we searched various databases including Springer, PubMed, ACM Digital Library, EBSCOHost, ProQuest, Google Scholar, Elsevier Scopus, and IEEE Xplore. The databases were selected to ensure that articles across various fields would be accessed for the study.

As shown in **Table 2A**, various keywords were used in the search process such as "Physical Activity," "Physical Activity Applications or Apps," "Sedentary Behavior or Behaviour," "Sedentary Behavior or Behaviour Applications or Apps," "Sedentary Lifestyle," "Prolonged Sedentary," "Prolonged Sedentary Behavior," "Prolonged Sedentary Sitting," "Prolonged Sitting," "Physical Activity and Sedentary Behavior," "Persuasive Technology and Physical Activity," "Persuasive Technology and Sedentary Behavior," "Persuasive Technology and Physical Activity and Sedentary Behavior," "Persuasive Technology Exercise," "Persuasive Technology Fitness," "Physical Activity and Gamification," "Physical Activity and Exergames," "Exercise Applications Or Apps," "Fitness Applications or Apps," "Exergames or Mobile Exergames." The search was refined through the use of Boolean terms such as "Persuasive Technology AND Physical Activity AND Sedentary Behavior." We adapted **Table 2A** from the previous work done by Wang et al. [7] and refined using more keywords identified from the literature, in the refine process.

The search in the databases was also refined using an inclusion and exclusion criteria. The first criteria was to include recent articles, so those articles published earlier than the year 2003 were excluded from the search because the first paper in the field of persuasive technology was introduced by Fogg [32] as a seminar paper in the year of 2002. Accordingly, most papers in the area of PT where published from the year of 2003. This was also to ensure that the findings reported in the studies were current and not outdated. The second criteria was that only articles that were in English were selected for the study. The search was run through the databases to locate relevant articles. The reference lists of these articles were also reviewed to further identify other potentially relevant articles.

3.1. Analysis and Coding Scheme

We retrieved 1,393 articles, of which 1,077 articles were identified through database searching, and 316 articles were identified through reviewing the reference lists of the obtained articles. There were 637 duplicate articles excluded from the total of 1,393 articles. The titles of these articles were examined, and those found not to be suitable were excluded, such as those that targeted health domains other than PA/SB. Overall, we identified 756 unique titles, of which 338 articles were excluded by titles, and after evaluating the abstracts of the remaining 418 articles, 170 articles were selected for final analysis. The study identification process is summarized in **Figure 1** [a PRISMA flow diagram [33]].

In the second step of this review, we coded the articles by creating an excel coding sheet for the PT analysis. As a starting point, we adopted a coding sheet that was developed and

TABLE 2 | (A) Search terms and combination methodology for articles selections; **(B)** Persuasive technology classifications and coding scheme analysis—Adapted from Orji and Moffatt [4].

(A)

Search terms method

Numbers	Terms	Combinations		Search terms
1	Physical	1 and 2	1 and 2 and 7	○ Physical Activity ○ Physical Activity applications or apps ○ Sedentary Behavior or Behavior ○ Sedentary Behavior or Behavior applications or apps ○ Sedentary Lifestyle ○ Prolonged sedentary ○ Prolonged sedentary behavior ○ Prolonged sedentary sitting ○ Prolonged sitting ○ Physical activity and sedentary behavior ○ Persuasive Technology and Physical activity ○ Persuasive Technology and sedentary behavior ○ Persuasive Technology and Physical activity and sedentary behavior ○ Persuasive Technology Exercise ○ Persuasive Technology Fitness ○ Physical activity and Gamification ○ Physical activity and Exergames ○ Exercise applications or apps ○ Fitness applications or apps ○ Exergames or Mobile exergames ○ Fitness Technology ○ Exergame Technology ○ Fitness ○ Exergames
2	Activity	1 and 2 and 3 and 6	1 and 2 and 3 and 6 and 4 and 5	
3	Sedentary	1 and 2 and 11	1 and 2 and 12	
4	Persuasive	1 and 2 and 13	1 and 2 and 14	
5	Technology	1 and 14	1 and 15	
6	Behavior or behavior	1 and 2 and 15	3 and 6	
7	Applications or apps	3 and 6 and 7	9 and 3	
8	Lifestyle	9 and 3 and 6	9 and 3 and 7	
9	Prolonged	9 and 3 and 10	9 and 10	
10	Sitting	4 and 5 and 1 and 2	4 and 5 and 3 and 6	
11	Exercise	4 and 5 and 1 and 2 and 3 and 6	4 and 5 and 11	
12	Fitness	4 and 5 and 12	15	
13	Gamification	16 and 15	11 and 7	
14	Games	16 and 1 and 2 and 14	12 and 7	
15	Exergames	3 and 8	12 and 5	
16	Mobile	12	15 and 5	

(B)

PT classifications and coding scheme

S/N	Identification	Examples/meaning
1	Papers	Name of the research papers and articles.
2	Author(s)	Name of the author(s) who wrote a research paper and conducted a study.
3	Year	The year of when the study was conducted.
4	Domain Focus	PA, SB, Eating, Smoking, Stress, Obesity, Sitting Postures, Mental Health, etc.
5	Technology	Mobile, Web, Games, Computer applications, Ambient displays, etc.
6	Evaluation Methods	Quantitative, Qualitative, and Mixed.

(Continued)

TABLE 2 | Continued

S/N	Identification	Examples/meaning
7	Persuasive Strategies	Motivational affordance strategies used in a PT system design.
8	Duration of Evaluation	Minutes, Hours, Days, Weeks, Months, and Years.
9	Behavior Theories	Theories used in a PT system design or assessment.
10	Targeted Outcomes	Behavior, Attitudes, Awareness, Adherence, Motivation, Feasibility, Cognitive, etc.
11	Targeted Audience based on their Age Group	Children, Teenagers, Young Adults, Adults, Elderly, etc.
12	Targeted Audience based on their Status/Occupations/Health Conditions	School students, University students, Office workers, Overweight and Obese, Nurses, Patients with type 2 diabetes, etc.
13	Number of Participants	Number of participants who participate in the assessment of a study.
14	Venue	CHI, UbiComp, Persuasive, MobileHCI, Pervasive Health, Health Informatics, JMIR, etc.
15	Effectiveness/Evaluation Outcomes	Identifying whether the study was successful or not successful.
16	Country/Region of a Study	Country or region where the study was conducted.

validated by Orji and Moffatt [4] and refined it by adding new coding categories that emerged as we iteratively analyzed our data. **Table 2B** shows how we classified and coded the articles. Once the articles were identified, they were coded and classified, as shown in **Appendix 1**.

4. RESULTS

The analysis of PTs for physical activity and SB reveal some interesting findings, as shown below. The findings are presented under various categories such as the year and country in which the technology was developed, the platforms, behavioral and psychological outcomes targeted, and the evaluation results of the PTs. The summaries of all the reviewed papers are as shown in **Appendix 1**. For the papers that have more than one study, we combined the findings for all the studies in the paper. For example, we reported the total number of participants, all the persuasive strategies used, and the total duration of all the studies in each paper.

4.1. Persuasive Technology for Physical Activity and Sedentary Behavior by Year and Country

As shown in **Table 3A** and **Figure 2A**, a large number of articles and studies were published after compared to before 2011. In recent years, there has been a sharp increase in the number of articles published since 2012, although the number fluctuated year to year from 2012 to 2019. It is important to mention that while 2019 appears to have the lowest number of studies since 2012, this is probably because most papers for 2019 are yet to be published at the time of this study, third quarter of 2019.

As it is evident from **Table 3B** and **Figure 2B**, the studies were conducted in 29 different countries, with most of the studies coming from the USA, 56 (33%). This is followed by the UK with a total of 16 articles. Australia and the Netherlands are in third place, with a total of 15 articles for each. Canada and Germany are in fourth with a total of 9 articles. Only one article did not specify the country where the study was conducted, only mentioning the continent such as North America, Europe, and Asia.

4.2. Effectiveness of Persuasive Technology for PA and SB

Table 4 and **Figure 3** show a summary of the results of the effectiveness of PT for PA and SB reviewed in this paper. We found that 87 (51%) studies reported fully successful outcomes, and 50 (29%) studies reported partially successful outcomes from using the PT to achieve desired behaviors and attitudes related to PA and/or SB. Partially positive results are used to describe studies that reported a combination of positive with negative or no effect results [4]. However, only 4 (2%) of the studies reported completely unsuccessful results. In the studies reviewed, 6 (4%) did not specify the outcomes of the technology, and 23 (14%) of the articles did not evaluate their PT design. As a result, most of the reviewed studies (80%) reported successful outcomes, whether fully or partially, while only 4% of the studies were

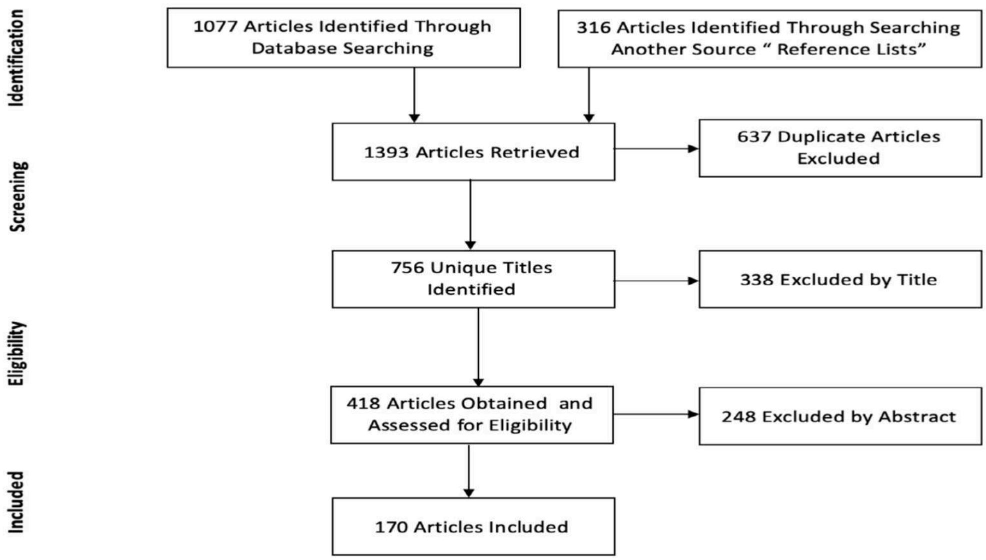


FIGURE 1 | PRISMA flow diagram for the study selection workflow.

unsuccessful. This means that PTs are effective tools to persuade people in practicing more PA and reducing their SB.

4.3. Major Technology Platforms Employed in PTs for Physical Activity and Sedentary Behavior and the Effectiveness of PTs

Figure 4A provides a summary of the major technology platforms employed to design the PTs for PA and SB. Mobile and handheld devices were the most used platform with a total of 61 studies (36%), followed by platforms that employed games and gamifications with precisely 33 (19%) studies, as well as web and social networks that placed second with 32 studies (19%). The games category includes all the interventions that were delivered in the form of games, irrespective of whether the game is web-based, a mobile, or a desktop device. We found that 31 (18%) studies used commercially available sensors and other activity trackers (e.g., Fitbit, Pebble smartwatch, ActivPAL, and ActiGraph), whereas 19 (11%) used custom-designed sensors and activity trackers that have been designed by the researchers in their studies. Ambient and public displays came in fifth place with 16 (9%) studies using this platform, this was followed by the interactive workstations and chairs with just 12 (7%) studies. Computer-based platforms such as desktop and laptop were the least frequently employed platform for delivering PTs for physical activity and sedentary behavior, with only 10 (6%) studies.

It is important to mention that most of the reviewed studies employed more than one technology platform in their PT design. Generally, the second most employed technology platforms after the mobile and handheld devices are activity trackers and sensors (whether commercial or custom-designed) with a total of 42 studies (29%). Consequently, by considering the use of embedded sensors in mobile devices, we can notice that the dominant technology platforms employed in the PTs for PA and SB were activity trackers and sensors and most PT employing them were successful. Thus, it is essential to employ activity trackers and sensors in the PT design to track users' performance and to provide them with accurate feedback about their activity progress to motivate them to change their unhealthy habits such as SB.

Figure 4B demonstrates the effectiveness of employing PT with regards to the technology platforms. For the mobile and handheld devices, we found that 48 (79%) of the studies reported successful results; that is, studies with partially successful and those with fully successful results. Precisely, 28 (58%) studies were fully successful, and 20 (42%) studies were partially successful. For the games, out of 33 studies employing them, 19 (58%) showed fully successful outcomes, 7 (21%) displayed partially successful outcomes, just 1 (3%) reported unsuccessful outcomes, and 6 (18%) did not provide evaluations. For the commercially available sensors and activity trackers, out of 33 studies using them, 12 (36%) reported fully successful results, 14 (43%) showed partially successful results, only 1 (3%) reported unsuccessful outcomes, 3 (9%) reported unspecified results, and 3 (9%) did not evaluate their studies. For the websites and social networks, out of 31 studies implementing them, 16 (52%) reported fully successful results, 9 (29%) showed partially successful results, only 1 (3%) did not specify the results, and 5 (16%) did not evaluate their PTs. For the custom made sensors and activity trackers, out of 19 studies designed them, 10 (53%) reported fully successful results, 4 (21%) provided partially successful results, 4 (21%) did not show evaluations, and only 1 (5%) reported unspecified results. For the ambient and public displays, out of 16 studies employing them, 9 (56%) reported fully successful results, 4 (25%) showed partially successful results, 2 (13%) reported unsuccessful outcomes, and 1 (6%) did not evaluate their studies. For the interactive workstations and chairs, out of 12 studies implementing them, 8 (67%) reported fully successful results, 1 (8%) showed partially successful results, only

TABLE 3 | (A) Persuasive technology for physical activity and sedentary behavior trends by year; **(B)** Persuasive technology for physical activity and sedentary behavior by study country/region.

(A)

Counrty	Study	Total	Overall of % 170
2003	[34]	1	1%
2004	[35]	1	1%
2005	[36]	1	1%
2006	[37, 38, 39, 40, 41]	5	3%
2007	[42, 43]	2	1%
2008	[44, 45, 46, 47, 48, 49, 50, 51]	8	5%
2009	[52, 53, 54, 55, 56, 57, 58, 59]	8	5%
2010	[60, 61, 62, 63, 64, 65, 66, 67, 68]	9	5%
2011	[69, 70, 71]	3	2%
2012	[72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91]	20	12%
2013	[92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113]	22	13%
2014	[114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131]	16	9%
2015	[132, 133, 134, 130, 135, 136, 131, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146]	17	10%
2016	[147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165]	19	11%
2017	[166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176]	11	6%
2018	([177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194]; [195])	19	11%
2019	[196, 197, 198, 199, 200, 201, 202, 203]	8	5%

(B)

Counrty	Study	Total	Overall of % 170
USA	[34, 35, 39, 41, 51, 44, 46, 50, 54, 59, 60, 64, 68, 69, 74, 80, 84, 85, 87, 89, 90, 93, 96, 99, 100, 102, 105, 106, 114, 117, 123, 125, 127, 128, 134, 136, 131, 140, 148, 150, 156, 158, 162, 164, 165, 166, 167, 173, 184, 188, 189, 190, 191, 199, 200, 203]	56	33%
Australia	[61, 75, 73, 92, 94, 96, 151, 98, 147, 152, 174, 201] [58, 109, 126]	15	9%
Austria	[49, 81, 88, 111, 141]	5	3%
Portugal	[120, 135, 186, 139, 159]	5	3%
Canada	[67, 72, 77, 83, 96, 122, 185, 187, 199]	9	5%
UK	[40, 42, 63, 78, 118, 79, 82, 101, 104, 124, 142, 154, 161, 168, 171, 196]	16	9%
Russia	[169]	1	1%
Malaysia	[157]	1	1%
Israel	[129]	1	1%
Thailand	[137]	1	1%
Switzerland	[38, 115, 130, 198]	4	2%
Germany	([62, 95, 170, 176, 177, 178, 192, 193]; [202])	9	5%
Netherlands	[37, 52, 53, 57, 66, 70, 71, 91, 110, 113, 195, 145, 160, 175, 181]	15	9%
United Arab Emirates (UAE)	[103]	1	1%
Taiwan	[116]	1	1%
Italy	[76, 119, 132]	3	2%
Finland	[55, 133, 153]	3	2%
Mexico	[48, 65, 112]	3	2%
South Korea	[43, 86, 108, 138, 155]	5	3%
Ireland	[34, 96, 97, 163]	4	2%
Belgium	[143, 144, 149, 197]	4	2%
France	[183]	1	1%
Norway	[182]	1	1%
Singapore	[56]	1	1%
Brazil	[45]	1	1%
China	[172]	1	1%
Japan	[36, 47, 107]	3	2%
Nigeria	[199]	1	1%
Spain	[146, 179, 180, 194]	4	2%
North America, Europe, Asia	[121]	1	1%

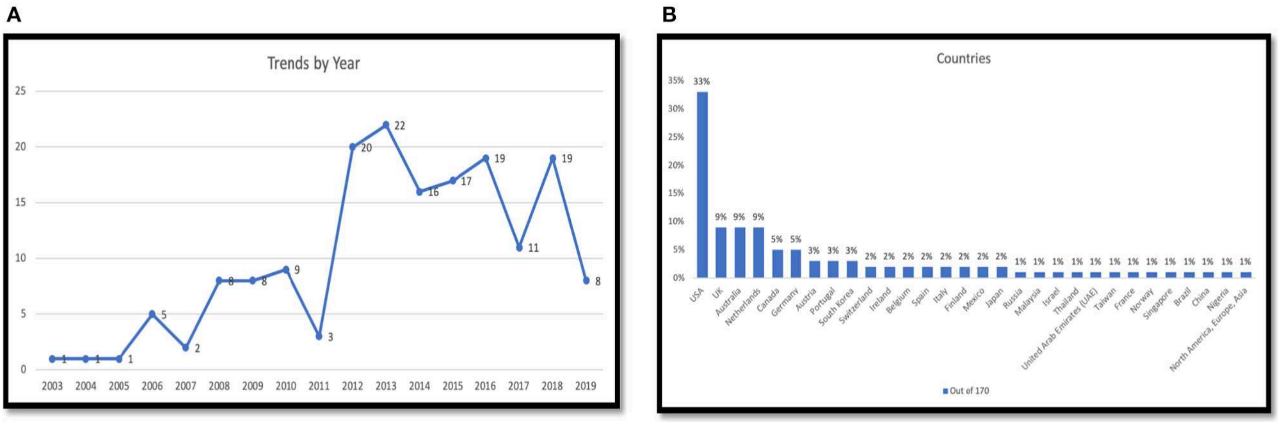


FIGURE 2 | (A) Persuasive technology for physical activity and sedentary behavior trend by year; **(B)** Persuasive technology for physical activity and sedentary behavior by study country/region.

TABLE 4 | Summary results of Persuasive Technology (PT) Effectiveness in Physical Activity (PA) and Sedentary Behavior (SB).

Results	Study	Total	Overall of % 170
Successful	[34, 58, 35, 37, 38, 39, 70, 41, 44, 45, 46, 47, 48, 49, 50, 54, 56, 57, 60, 61, 75, 62, 63, 64, 66, 67, 68, 69, 72, 73, 76, 79, 80, 83, 85, 86, 89, 92, 93, 94, 95, 100, 101, 103, 109, 110, 114, 115, 116, 117, 118, 121, 122, 124, 126, 128, 130, 136, 131, 137, 139, 141, 144, 147, 148, 149, 150, 151, 153, 154, 155, 156, 158, 146, 163, 166, 169, 173, 174, 175, 176, 184, 188, 191, 196, 197, 203]	87	51%
Partially successful	([36, 42, 43, 51, 52, 53, 59, 71, 77, 81, 84, 96, 99, 104, 105, 112, 113, 120, 123, 125, 129, 132, 133, 134, 135, 186, 140, 142, 143, 160, 164, 167, 170, 172, 177, 178, 179, 180, 181, 182, 183, 185, 187, 190, 192, 193]; [195, 199, 201, 202])	50	29%
Unsuccessful	[91, 107, 108, 165]	4	2%
Unspecified	[40, 55, 97, 98, 138, 152]	6	4%
No evaluation	[65, 74, 78, 82, 87, 88, 111, 90, 102, 106, 119, 127, 145, 150, 159, 161, 162, 168, 171, 189, 194, 198, 200]	23	14%

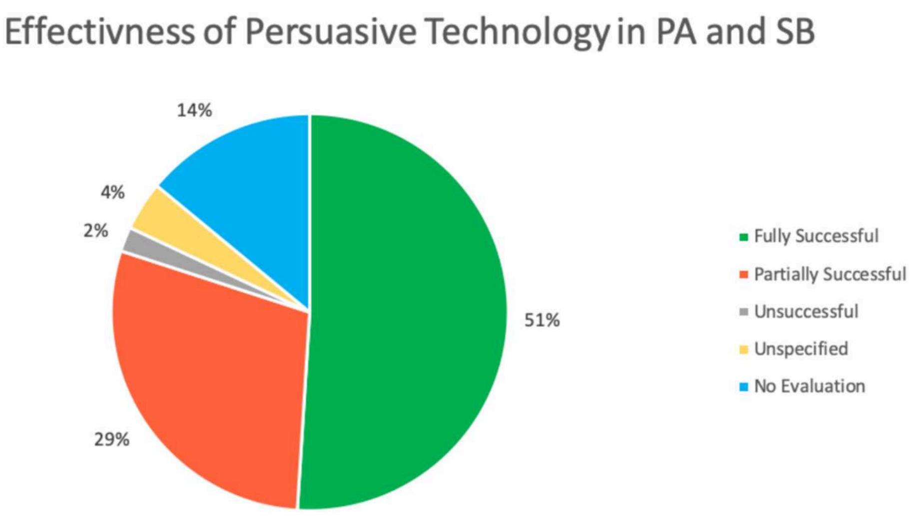


FIGURE 3 | Effectiveness of persuasive technology in physical activity and sedentary behavior.

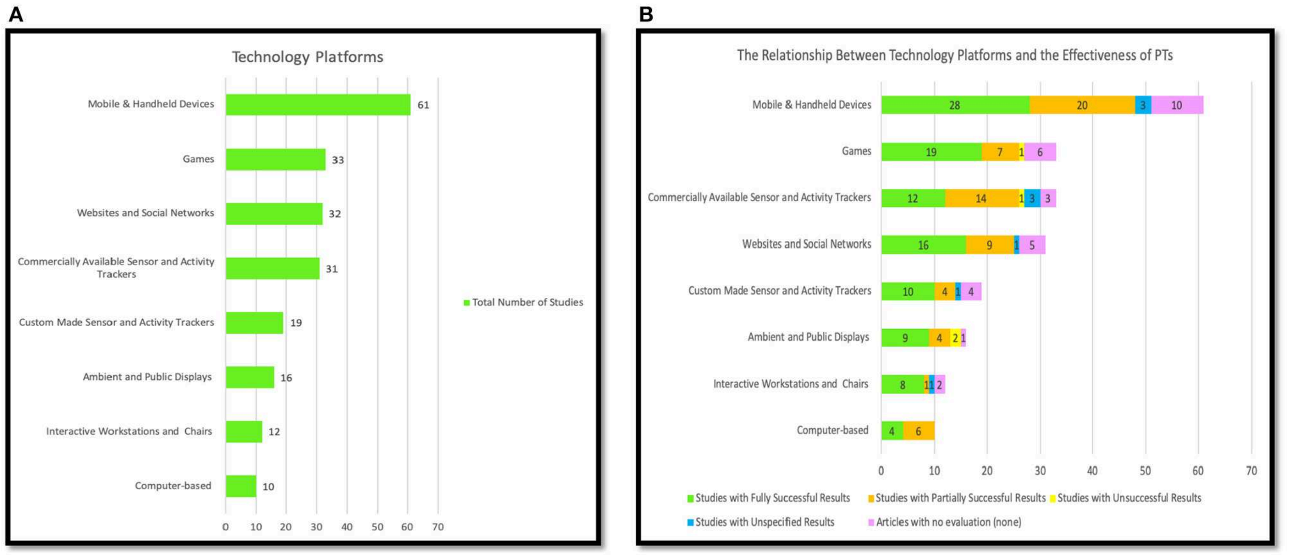


FIGURE 4 | (A) Persuasive technology platforms. **(B)** Technology platforms and the effectiveness of PTs.

1 (8%) did not specify the results, and 2 (17%) did not evaluate their PTs. For the computer-based technology such as a desktop, we found that 10 of the studies reported successful results; that were 6 (60%) studies with partially successful, and 4 (40%) studies with fully successful results. Overall, the findings show that the most effective technology platforms are mobile and handheld devices with 48 successful studies (whether fully or partially successful), followed by activity trackers and sensors (whether commercial or custom-designed) with 40 successful studies, and then games with 26 successful studies, followed by websites and SNSs with 25 successful studies.

4.4. Persuasive Strategies and Motivational Affordances

Table 5 and **Figure 5** show the strategies most commonly employed to bring about the intended behavioral outcomes in the PA and/or SB domains. Tracking and self-monitoring were the most frequently employed strategies with a total of 153 (90%) studies. Reminder ranked as the second most employed strategy with 72 (42%) studies, and personalization is the third most employed strategy with a total of 64 (38%) studies. Rewards and goal-setting ranked as the fourth and fifth frequently employed strategies with 54 (32%) studies and 53 (31%) studies respectively using the strategy. Other social support strategies (which refer to those strategies that did not belong precisely to the PSD model or those that were not specified such as social comments, tags, likes, chatting, and sending invitations, etc.) came sixth, with a total of 43 (25%) studies implementing these strategies. Simulation came in seventh place with a total of 42 (25%) studies, and praise came eight, with a total of 38 (22%) studies. Thirty-two (19%) studies employed the reduction strategy, which was the ninth most frequently used strategy. Suggestion and social competition strategies emerged as the tenth and eleventh most frequently used strategies with 30 (18%) studies employing each of them. Finally, tailoring, tunneling, social cooperation, surface credibility, social comparison, liking, and expertise credibility emerged as the 12, 13, 14, 15, 16, 17, and 18th most frequently used strategies, respectively, with a total of 29 (17%), 25 (15%), 19 (11%), 18 (11%), 17 (10%), 14 (8%), and 13 (8%) studies, see **Figure 5**.

4.5. Examples of Persuasive Strategies Employed in the Reviewed Studies

It is important to note that most of the studies employed more than one strategy at a time, and each strategy may have been implemented differently from one study to another. For example, the main strategy used by a study could be self-monitoring, but the app may also provide feedback that may appear in different formats such as audio, visual or textual feedback. It is essential to mention that we relied mainly on the PSD model [3] in sorting and organizing the persuasive strategies we obtained from the reviewed articles. **Table 1** summarizes the PSD model principles' "strategies." However, we identified some strategies that were not capured in the PSD model such as goal setting, punishments, self-report, and other social support strategies. For instance, goal setting is not part of the strategies highlighted in the PSD model; however, it is clearly an example of the persuasive strategies that have been employed in many PA and SB applications.

Other social support strategies (which refer to strategies that did not belong precisely to the PSD model or those that were not specified) such as (social sharing, social set/accept challenges, social posting feeds, social sending likes, social follow, social messages exchange (e.g., sending encouraging feedback, invitation, chatting), social interaction (e.g., communicating via video-conferencing "video streams, microphone"), social giving comments, and tagging).

4.5.1. Punishment Strategy

The punishment strategy also known as "negative reinforcement" does not belong to any strategy in the PSD model. An example

TABLE 5 | Persuasive strategies for PT of physical activity and sedentary behavior.

#	Motivational strategies/ Affordances	Studies with fully successful results	Studies with partially successful results	Studies with unsuccessful results	Studies with unspecified results	Articles with no evaluation (none)	Total number of studies	Average out of % 170 for each
1	Reduction	[41, 45, 60, 61, 64, 67, 70, 76, 80, 86, 100, 116, 117, 139, 157, 166, 175, 191, 196]	[104, 123, 135, 140, 181, 182, 185]		[55]	[74, 106, 150, 159, 198]	32	19%
2	Tunneling	[37, 45, 64, 66, 86, 94, 100, 117, 122, 137, 139, 157, 173, 191, 197]	[52, 104, 123, 182, 192]			[102, 111, 119, 127, 189]	25	15%
3	Tailoring	[37, 50, 54, 73, 79, 85, 89, 98, 100, 126, 130, 149, 157, 146, 174, 175, 197]	[36, 81, 112, 123, 183, 185, 199]		[55, 98]	[88, 119, 168, 198]	29	17%
4	Personalization	[41, 60, 63, 67, 68, 69, 70, 72, 73, 76, 83, 85, 86, 89, 98, 114, 124, 126, 130, 137, 141, 149, 153, 157, 158, 169, 174, 175, 191, 196, 197, 203]	[42, 43, 51, 71, 123, 167, 170, 178, 182, 183, 185, 186, 187, 190, 193, 195, 199, 202]	[91, 107]	[55, 97, 98]	[88, 90, 106, 127, 150, 159, 162, 168, 189, 200]	64	38%
5	Tracking/Self-monitoring	[34, 35, 37, 38, 39, 70, 41, 44, 45, 46, 47, 48, 50, 57, 60, 61, 75, 62, 63, 64, 67, 68, 69, 73, 76, 79, 80, 85, 86, 89, 92, 93, 94, 95, 100, 101, 103, 109, 114, 115, 116, 118, 121, 124, 126, 128, 130, 136, 131, 137, 139, 141, 144, 148, 149, 151, 153, 154, 155, 156, 157, 158, 146, 163, 166, 169, 173, 175, 176, 184, 188, 191, 196, 197, 203]	([36, 42, 43, 51, 52, 53, 59, 77, 81, 84, 96, 99, 104, 105, 112, 113, 120, 123, 125, 129, 132, 133, 134, 135, 186, 140, 142, 143, 160, 167, 170, 172, 177, 178, 179, 180, 182, 183, 185, 187, 192, 193]; [195, 199, 201, 202])	[91, 107, 108, 165]	[40, 55, 97, 98, 138, 152],	[74, 78, 119, 150, 168, 198] [65, 82, 87, 90, 102, 106, 111, 127, 145, 159, 161, 162, 171, 189, 194, 200]	153	90%
6	Simulation	[34, 58, 35, 39, 46, 56, 64, 66, 67, 72, 73, 76, 83, 89, 98, 122, 124, 126, 130, 141, 148, 157, 174, 175, 176]	[36, 104, 120, 134, 160, 180, 181]	[91, 107]	[98, 138]	[82, 88, 111, 106, 119, 194, 198]	42	25%
7	Rehearsal	[157]					1	1%
8	Praise	[37, 41, 44, 46, 48, 50, 64, 68, 85, 92, 100, 114, 157, 146, 191, 196, 197]	[51, 81, 112, 113, 195, 129, 134, 135, 177, 179, 183, 192, 193]	[91]	[55]	[74, 87, 90, 106, 119, 189]	38	22%
9	Rewards	[39, 44, 49, 60, 61, 75, 64, 66, 68, 69, 76, 80, 85, 110, 114, 115, 116, 121, 122, 124, 137, 141, 153, 191, 196]	([43, 52, 81, 84, 104, 112, 129, 135, 167, 177, 178, 181, 182, 185, 192, 193]; [199])	[91, 107]	[55]	[65, 74, 82, 102, 106, 145, 161, 171, 189]	54	32%
10	Punishments	[39, 49]	[43]	[107]		[119]	5	3%
11	Reminders	[35, 38, 44, 50, 54, 66, 69, 70, 79, 85, 89, 92, 93, 95, 114, 126, 130, 136, 141, 144, 151, 153, 154, 155, 157, 158, 169, 175, 176, 191, 196, 203]	[43, 51, 53, 71, 81, 84, 96, 99, 112, 113, 195, 120, 123, 133, 140, 160, 164, 167, 170, 178, 179, 183, 185, 186, 187, 190, 202]	[107, 165]	[55, 138]	[74, 87, 102, 111, 145, 150, 171, 189, 194]	72	42%
12	Suggestion	[45, 47, 54, 66, 70, 85, 100, 117, 122, 126, 136, 131, 139, 149, 157, 146, 173, 175]	[81, 84, 96, 112, 120, 123, 135, 195]		[138]	[87, 150, 168]	30	18%
13	Similarity	[157]	[178]				2	1%
14	Liking	[72, 75, 124, 141, 175]	[36, 123, 181, 185, 193]	[91, 108]		[189, 198]	14	8%
15	Social role	[175, 191]	[187]		[55]	[106, 119, 159]	7	4%
16	Trustworthiness	[100, 191]	[81]		[55]	[168]	5	3 %
17	Expertise	[100, 141, 144, 149, 157, 191, 196, 197]	[53, 81, 187]		[55]	[168]	13	8%
18	Surface credibility	[100, 131, 144, 158, 175]	[84, 105, 132, 135, 202]	[91]	[55]	[74, 90, 102, 119, 168, 171]	18	11%

(Continued)

TABLE 5 | Continued

#	Motivational strategies/ Affordances	Studies with fully successful results	Studies with partially successful results	Studies with unsuccessful results	Studies with unspecified results	Articles with no evaluation (none)	Total number of studies	Average out of % 170 for each
19	Real-world feel	[64, 100, 157, 196]					4	2%
20	Authority	[100, 191]	[71]			[150, 168]	5	3%
21	Third-party endorsements	[100, 157]				[168, 171]	4	2%
22	Verifiability	[100, 157]					2	1%
23	Social learning	[38, 69, 157]	[172, 199]			[168, 200],	7	4%
24	Social comparison	[157, 175, 203]	([42, 51, 113, 129, 143, 177, 178, 179]; [199])	[91]		[74, 82, 119, 200]	17	10%
25	Normative influence	[100, 157, 191]	([113, 177, 178])			[168]	7	4%
26	Social facilitation	[38, 157]	[81]				3	2%
27	Social cooperation	[39, 41, 47, 48, 68, 86, 115, 184, 191]	[43, 81, 172, 177, 180, 192, 199]			[65, 74, 119, 145, 162, 198]	19	11%
28	Social competition	[34, 58, 39, 41, 45, 46, 47, 56, 60, 66, 68, 86, 94, 103, 115, 116, 153, 157, 184]	[42, 43, 51, 134, 177, 181, 182]	[91]		[74, 145, 162]	30	18%
29	Social recognition & rankings	[153, 157]	[172]	[91]		[145, 168]	6	4%
30	Other social support strategies	[38, 39, 41, 47, 48, 58, 64, 66, 68, 69, 85, 86, 100, 103, 115, 121, 124, 137, 146, 188, 191, 197, 203]	[42, 51, 53, 84, 112, 113, 143, 160, 172, 179]	[91, 108]	[97]	[65, 74, 90, 145, 150, 168, 171]	43	25%
31	Goal setting	[38, 39, 44, 47, 48, 63, 64, 68, 85, 92, 100, 114, 121, 126, 139, 146, 175, 184, 188, 191, 196, 197]	[43, 51, 52, 53, 81, 84, 99, 104, 112, 113, 123, 129, 135, 186, 143, 179, 181, 182, 183, 185, 187, 199]		[55, 97]	[74, 82, 106, 168, 171, 189, 200]	53	31%
32	Feedback from users (Self-Report)	[50, 68, 70, 158]	[190]			[90]	6	4%

of a punishment strategy was a sad or an angry emotional facial expression of fish in a social computer game called "Fish 'n' Steps" [39], and a negative expression, such as in the "Persuasive Art" ambient mirror system [107]. It's also exemplified in apps where users lose some points for not meeting their goals.

4.5.2. Tracking/Self-Monitoring Strategy

The examples of tracking/self-monitoring strategy were diverse in the reviewed papers. For example, tracking/self-monitoring could be in the form of textual and visual feedback of a user's progress, step counts, approximate burnt calories, and goal completion as can be seen from **Figure 6** of the mobile activity tracker system called "Habito" [135] and the "On11" mobile system [123]. Real-time and vibration feedback were used as tracking/self-monitoring in the mobile game called "LocoSnake" because the phone vibrates when the snake's head goes near a piece of fruit [76]. Tracking/self-monitoring also was represented as a graphical and informative art visualization such as in the "Spark" web application [80]. Another tracking/self-monitoring strategy was the "Pediluma" show activity tracker device that monitors the wearer's PA by providing various light intensity levels regarding the user's status based on whether he/she was engaged in PA (e.g., walking) or sedentary [69]. The sculpture in the "Breakaway" ambient display system was used as a tracking/self-monitoring strategy and a reminder strategy since there was a connection between the user's movements (whether sedentary or physically active) and the sculpture placed on the office workers' desks [35].

4.5.3. Authority Strategy

An authority strategy was implemented as an example, as was presented in the "PRO-fit" system by using the OAuth 2.0 protocol [150]. The Calendar Integration Manager (CIM) module allows the "PRO-fit" system to integrate many calendar services providers such as Yahoo, Hotmail, Google, etc. [150].

4.5.4. Third-Party Strategy

A third-party endorsement strategy was used in the "WragaFit" application, as the PA goals were set by the Ministry of Health [157]. Another example was represented in the "WeightBit" application, which used the Apple Technology Company's Health Kit [171].

4.5.5. Simulation Strategy

The simulation strategy was found in PTs such as the mobile game called "LocoSnake," in which the user represents a virtual snake in the game. When the user walks and moves in the

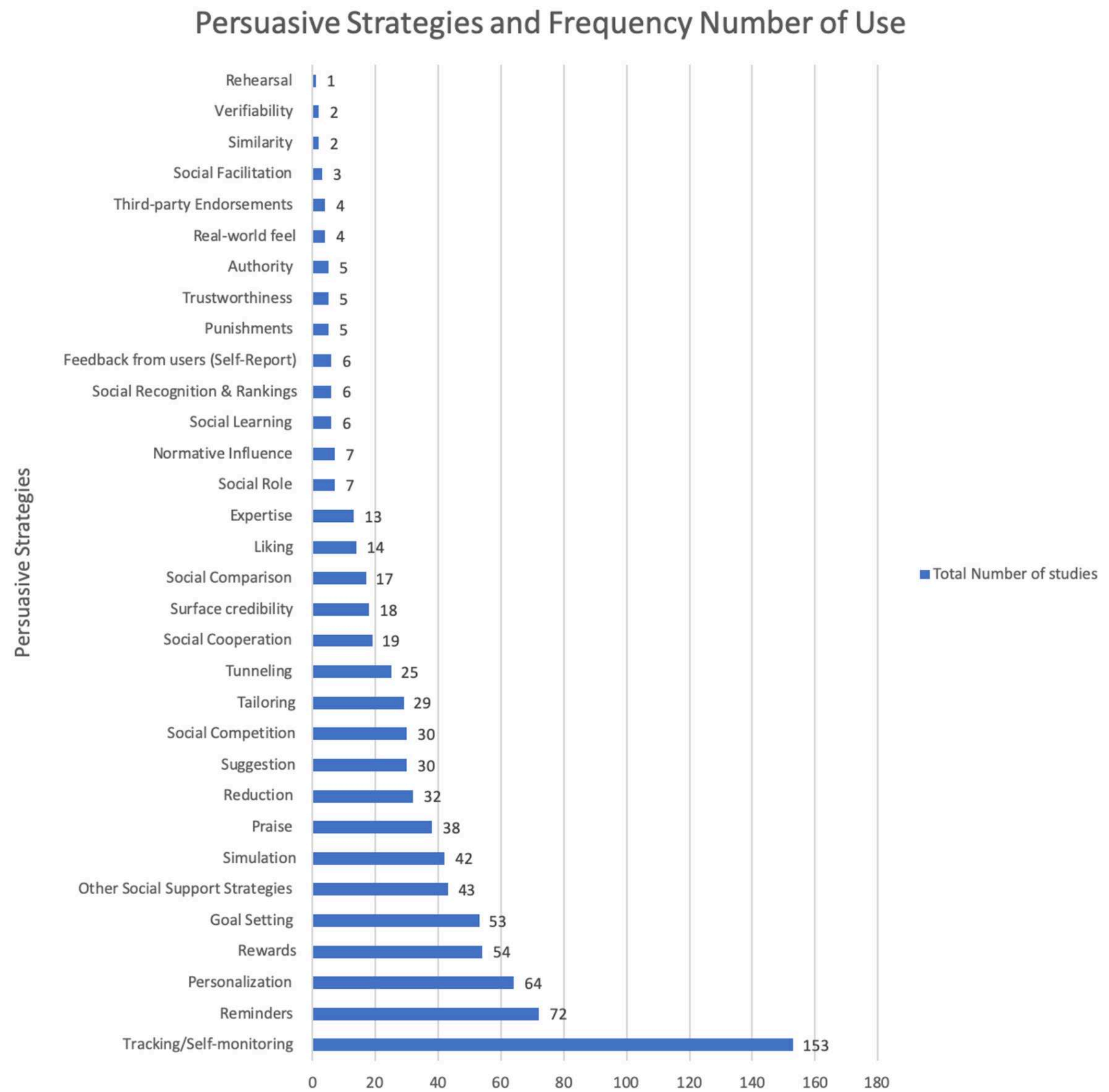


FIGURE 5 | Persuasive strategies and frequency of use.

real world, this controls the movement of the snake with the help of GPS and visualized satellite map technologies [76]. Another example was the interactive "GrabApple" game, which requires a player to make movements in the physical world such as raising their hands and jumping to pick up virtual falling green and red apples in the game on the screen [122]. Another instance was the "Energy Browser" system which allows users to wear activity sensor devices, and to observe the effects of their healthy physical movements while walking or running on treadmills [36]. Another example of a simulation strategy was in a web and smartphone game called "Phone Row" in which the users control the movements of a virtual boat through a virtual route on an outer screen [91]. The previously mentioned examples of a simulation strategy gave the user the ability to observe the connection between the cause and effect regarding his/her behavior, which reflects the definition of the simulation strategy in the PSD model [3].

4.5.6. Suggestion Strategy

The suggestion strategy or what is known as persuasive messages or recommendations were shown in the following example (e.g., "Try walking when talking on the phone. During your call with Bob, you were sedentary," "Last week, you reached your daily walking goal two times, try updating it to 8 km") [135].

4.5.7. Goal Setting Strategy

A goal setting strategy was used in the smartphone application called "On11," which allows users to set their performance goal to enable the system to recommend the users suitable activities

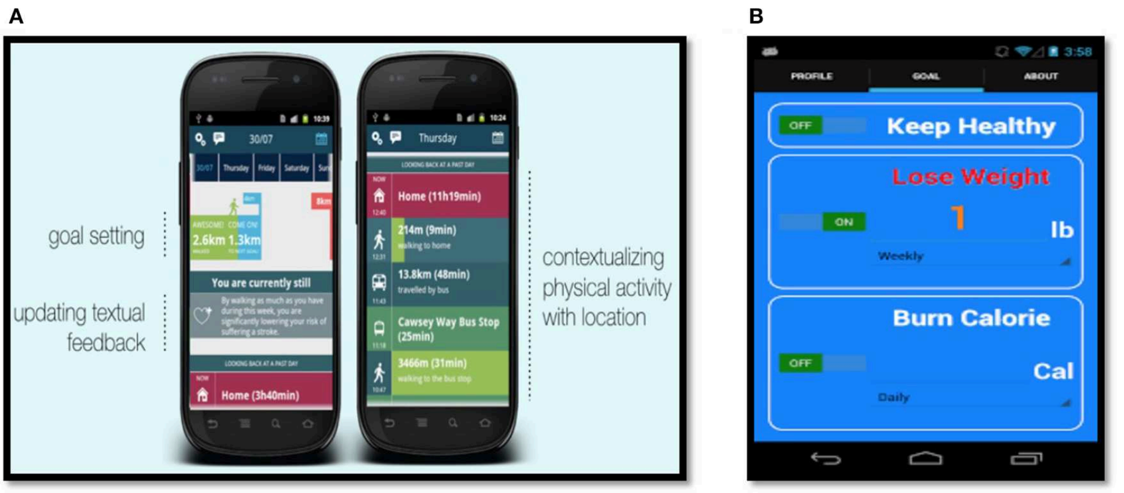


FIGURE 6 | (A) Visual and textual feedback in habito mobile activity tracker system [135]; **(B)** Visual and textual feedback in on11 mobile system [123].

based on their conditions (time, weather, location) to assist them to meet their goals [123], as shown in **Figure 7A**.

4.5.8. Tunneling Strategy

An example of a tunneling strategy was found in the "On11" system by generating walking routes to guide users through the use of Google Directions API [123]. The smartphone-based exergames called "Go Run Go" [137], and "BunnyBolt" [102] as shown in **Figures 7B,C**, represented a tunneling strategy that uses storyline scenarios to guide a player throughout the games.

4.5.9. Reward Strategy

The reward strategy was exemplified as badges in the "BunnyBolt" game [102]. A virtual trophy or stars were used in the "Polar FT60" system as rewards [55]. Intelligent musical stairs known as "Social Stairs" were implemented as a reward strategy by triggering music corresponding to the user's steps on stairs [110]. As shown in **Figure 7A**, there were seven visual growth levels for the virtual fish in the "Fish in Steps" desktop game and a happy facial expression of virtual fish was used as reward and tracking/self-monitoring strategies for users [39].

4.5.10. Priase Strategy

The praise strategy was used as an encouraging text message ("Keep walking! You can do it!") [85]. Another example of a praise strategy to motivate users to do more PA was shown in the heart rate monitor system called "Polar FT60" by delivering encouraging verbal feedback such as "Maximal performance improving," "Well done!," or "Excellent!" [55].

4.5.11. Tailoring Strategy

The tailoring strategy was employed in the mobile phone text messaging system [54] by providing information and tips on the PA and healthy eating domains tailored to African-American women who participated in a weight management program. Similarly, a tablet-based application called "Agile Life" was designed to be tailored to the elderly by giving them PA information chunks [81]. Another example of a tailoring strategy was used by micro-blogging sites like "Twitter," which was tailored to encourage teenage girls to exercise through the use of social media supports [68].

4.5.12. Reduction Strategy

The reduction strategy was used in different ways as represented in the reviewed articles (e.g., targeting simple behavior such as stretching and walking) as shown in the "WragaFit" application [157]. It was also seen in the "LocoSnake" game [76], as users could select the level of the game from three difficulty levels (easy, medium, and hard). In addition, a reduction strategy was represented, for example, in the "CrowdWalk" mobile application [139], since the application provides a list of a location-based "walking challenges" through the use of a map visualization to give the user an easy way to engage in nearby activities and challenges.

4.5.13. Social Comparison and Social Learning Strategies

The social comparison and social learning were used in the "WragaFit" smartphone application [157] as highlighted in **Figure 8A**. Another example of a social learning strategy was the "Pediluma" shoe activity tracker device that monitored the wearer's movements by providing varying intensities of a lighted cage [69].

4.5.14. Social Cooperation Strategy

The social cooperation was used in the tablet application "Agile Life" [81] to enable elderly users to engage with friends in PAs. Simulation and social comparison strategies were used as a mechanism on a group and individual level with the assistance of Facebook as in the "Active2Gether" system, so a user was able

FIGURE 7 | (A) On11 detour map [123]; **(B)** BunnyBolt game [102]; **(C)** BunnyBolt game scenario [102].

to compare his/her performance with others and notice the link between a cause and effect [175].

4.5.15. Social Competition and Social Recognition Strategies

Social competition and social recognition "ranking" strategies were used, such as in a Facebook application called "StepMarton" [63] that displays the entire number of steps for each user and his/her name in an order from the user with the highest number of steps on the top to the user with the smallest number of steps on the bottom.

4.5.16. A Real-World Feel Strategy

The example of a real-world feel strategy was shown in the "WragarFit" system by enabling users to accomplish each other's tasks on a "news feed" [157].

4.5.17. Social Facilitation Strategy

The social facilitation strategy was implemented in a mobile lifestyle coaching application [38] by allowing the achievements of an individual team member to be visible to the rest of the team and the achievement of the entire team to be visible to all members of a team and other teams.

4.5.18. Normative Influence Strategy

The normative influence strategy was employed in the "SitCoach" application since it stores the number of active minutes daily for each user and gives a notification for all users to observe the progress of each other [113].

4.5.19. Personalization Strategy

The personalization strategy was employed in the "StepMarton" application [63] by providing personalized Facebook notifications and in the "Alert Me" mobile application by delivering timely personalized messages to users and by allowing users to create personal profiles [169].

4.5.20. Self-Report Strategy

The self-report strategy was represented as feedback from a user to the system, such as in the "Time for Break" system [190], as a user provides feedback when responding to the reminders. Users in such a situation respond to the reminder question with

FIGURE 8 | **(A)** Social comparison and social learning in wragaFit application [157]; **(B)** Time for break system [190]; **(C)** Exerlean bike system [72].

either "Yes" or "No," and when choosing "Yes," the users enter a desirable duration for a break, and when they choose "No," the users type the reasons for not taking a break as shown in **Figure 8B**. The smart-watch application in the "ROAMM" monitoring system was used to collect self-reported data from users [158].

4.5.21. A Similarity Strategy

A similarity strategy was employed in the "WragaFit" system by providing images to older workers to make them feel familiar [157].

4.5.22. A Reminder Strategy

An example of a reminder strategy was mentioned in the "Time for Break" system [190] by issuing a textual notification as a question to a user (e.g., "You have been working for 30 min. How about taking a walk or standing up?") as shown in **Figure 8B**. Another instance of a reminder was represented in the "SitCoach" mobile application as an acoustic (buzzing) alert, a textual message, and a tactile reminder (vibration) [113]. A musical reminder in the "FLOW Pillow" system for the elderly was another way of implementing a reminder strategy [160].

4.5.23. Surface Credibility Strategy

The surface credibility strategy was employed in the "PersonA" system (a persuasive social network for PA) by providing security, confidentiality, and privacy features in the system [74]. Furthermore, a smartphone and web game known as "Phone Row" implemented a surface credibility strategy by offering a security mechanism through generating a new identifier for a present computer screen every time a user visits the webpage. In addition, a user was also required to scan a QR-code on the website [91].

4.5.24. Rehearsal Strategy

The rehearsal strategy was used by providing a video tutorial to educate users on appropriate techniques for doing stretching at the workplace [157].

4.5.25. Expertise Strategy

The expertise strategy was used by delivering healthy tips and information to older workers from an official medical source or fitness experts [157].

TABLE 6 | Comparative effectiveness of persuasive strategies of persuasive technology.

#	Motivational strategies/Affordances	Total number of studies with fully successful results	Total number of studies with partially successful results	Total number of studies with unsuccessful results	Total number of studies with unspecified results	Total number of articles with no evaluation (none)
1	Tracking/Self-monitoring	75	46	4	6	22
2	Reminders	32	27	2	2	9
3	Personalization	31	18	2	3	10
4	Simulation	24	7	2	2	7
5	Rewards	24	17	2	1	9
6	Other social support strategies	23	10	2	1	7
7	Goal setting	22	22	0	2	7
8	Reduction	19	7	0	1	5
9	Social competition	19	7	1	0	3
10	Suggestion	18	8	0	1	3
11	Praise	17	13	1	1	6
12	Tailoring	16	7	0	2	4
13	Tunneling	15	5	0	0	5
14	Social cooperation/Collaboration	9	7	0	0	6
15	Expertise	8	3	0	1	1
16	Liking	5	5	2	0	2
17	Surface credibility	5	5	1	1	6
18	Real-world feel	4	0	0	0	0
19	Feedback from users (Self-Report)	4	1	0	0	1
20	Social learning	3	2	0	0	2
21	Social comparison	3	9	1	0	4
22	Normative influence	3	3	0	0	1
23	Social role	2	0	0	1	3
24	Trustworthiness	2	1	0	1	1
25	Authority	2	1	0	0	2
26	Third-party endorsements	2	0	0	0	2
27	Verifiability	2	0	0	0	0
28	Social facilitation	2	1	0	0	0
29	Social recognition & Rankings	2	1	1	0	2
30	Rehearsal	1	0	0	0	0
31	Punishments	1	1	1	0	1
32	Similarity	1	0	0	0	0

4.5.26. Verifiability Strategy

The verifiability strategy was implemented in the "WragaFit" application, as users were able to verify the source of the provided health tips and information through an external link [157].

4.5.27. Trustworthiness Strategy

The trustworthiness strategy was implemented in the "Polar FT60" heart rate monitor because Polar is a trustworthy source of information [55].

4.5.28. Liking Strategy

The liking strategy was clearly shown in the Exerlean Bike System [72], as it provides children with attractive audial, textual, and visual representations for both Memory and ExerMath games and through the use of sensors and a stationary bike to easily enable the children to do PAs while responding to the games' assignments, as shown in **Figure 8C**. Other examples of liking and expertise strategies are found in the "Active2Gether" system when an expert designer was hired to design and provide recommendations on diverse aspects of the user interface to give the system an appropriate look and feel for the users [175].

4.5.29. Social Role Strategy

The reasoning engine feature in the "Active2Gether" system implemented a social role strategy by providing communication dialogue between the users and the system, which contains messages or questions for the users [175]. Moreover, a social role strategy was implemented as a "personal trainer" to guide

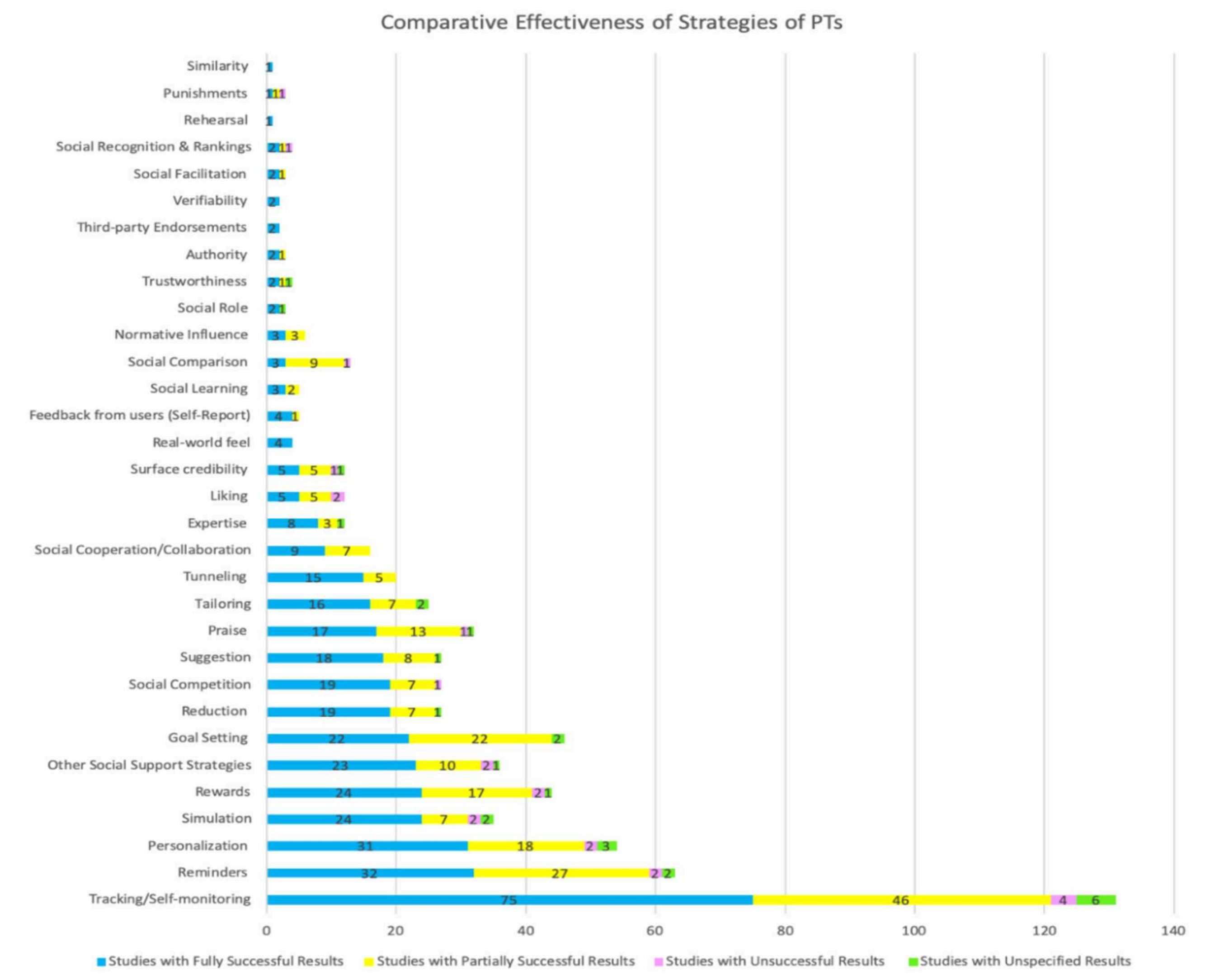


FIGURE 9 | Comparative effectiveness of persuasive strategies.

users' movements by giving verbal and personalized feedback as mentioned in the "Polar FT60" system [55].

4.6. Comparative Effectiveness by the Persuasive Strategies

Table 6 and **Figure 9** show the comparative effectiveness of PTs using persuasive strategies in the domain of PA and SB. The table and figure indicate that some strategies were applied more frequently and some tend to be more effective than others. For example, the tracking and self-monitoring strategy was employed in 153 (19%) studies, with a total of 75 (49%) studies reporting fully successful outcomes, 46 (30%) studies reporting partially successful outcomes, four (2%) studies reporting unsuccessful outcomes and six (4%) studies not specifying their outcomes, while 23 (15%) studies did not evaluate their strategies.

In summary, we reported the top twelve persuasive strategies most frequently used in the domain of PA and SB with respect to their effectiveness. As represented in **Table 6** and **Figure 9**, out of the total studies that implemented each persuasive strategy (see Section 4.4), tracking and self-monitoring ranked first with a total of 121 (79%) successful outcomes, followed by reminder and personalization, which ranked second and third with 59 (82%), and 58 (91%) successful results, respectively. Goal-setting came at fourth with 44 (83%) successful outcomes. Rewards ranked at fifth with 41 (76%) successful results. Other social support strategies ranked 6th with total numbers of 33 (77%) successful studies. Simulation and praise were at 7th and 8th with 31 (84%) an 30 (79%) successful studies, respectively. Reduction, social competition, and suggestion came in the 9th place with total numbers of 26 (81%) successful studies for each. Tailoring, tunneling, and expertise ranked 10, 11, 12th with 23 (79%), 20 (80%), and 11(85%) successful studies, respectively.

Generally, we noticed that the five most effective persuasive strategies employed were tracking/self-monitoring, reminders,

TABLE 7 | Behavior theories used in persuasive technology design.

Theories	Study	Total number of studies	Average out of % 170 for each
Transtheoretical model (TTM)	[39, 42, 44, 69, 82, 84, 85, 94, 135, 175, 191]	11	6%
Goal setting theory (GST)	[84, 107, 135, 186, 197]	5	3%
Theory of planned behavior (TPB)	[60, 74, 149, 161]	4	2%
Social cognitive theory (SCT)	[42, 74, 82, 94, 116, 126, 175, 191, 199, 203]	10	6%
Theory-driven design strategies (TDDS)	[80, 95, 149]	3	2%
Model-based reasoning (MBR)	[175]	1	1%
Dynamic computational model (DCM)	[175]	1	1%
Self-regulation theory (SRT)	[149, 175]	2	1%
Health action process approach (HAPA)	[175]	1	1%
Theory of reasoned action (TRA)	[103]	1	1%
Theory of meaning behavior (TMB)	[50, 60]	2	1%
Personality theory (PT)	[60]	1	1%
Theoretical domain framework (TDF)	[154]	1	1%
Self-determination theory (SDT)	[50, 57, 94, 143, 149, 153, 178]	7	4%
Unified theory of acceptance and use of technology (UTAUT)	[144]	1	1%
Grounded theory (GT)	[58, 172]	2	1%
Social production function (SPF) theory	[56]	1	1%
Cognitive dissonance theory (CDT)	[112]	1	1%
Theory of synchronization (TS)	[108]	1	1%
Wellness motivation theory (WMT)	[106]	1	1%
User-specific strategies (USS)	[106]	1	1%
Theoretical design principles (TDP)	[106]	1	1%
Contemporary psychology theory (CPT)	[132]	1	1%
Locomotor respiratory coupling (LRC) theory	[131]	1	1%
Hidden markov models (HMM)	[42]	1	1%
Theory of self-efficacy (TSE)	[82, 104]	2	1%
Social participation (SP)	[82]	1	1%
Classic learning theory (CLT)	[52]	1	1%
Operant conditioning theory (OCT)	[75, 107]	2	1%
Theory of Premack's principle (TPP)	[75]	1	1%
Regulatory focus theory (RFT)	[183]	1	1%
Flow theory (FT)	[160]	1	1%

(Continued)

TABLE 7 | Continued

Theories	Study	Total number of studies	Average out of % 170 for each
Unspecified (none)	[34, 35, 36, 37, 38, 40, 41, 51, 43, 45, 46, 47, 48, 49] [53, 54, 55, 59, 61, 62, 63, 64, 65, 66, 67, 68, 70, 71, 72, 73, 76, 77, 78] [79, 81, 83, 86, 87, 88, 111, 89, 90, 91, 92, 93, 96, 151, 97, 98, 99, 118] [100, 101, 102, 105, 109, 110, 113, 114, 115, 117, 119, 120, 121, 122, 123, 124, 125, 127] [128, 129, 133, 134, 130, 136, 137, 138, 139, 140, 141, 142, 145, 147, 148, 150, 152, 174] [155, 156, 157, 158, 159, 146, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 173, 176, 179] [177, 180, 181, 182, 184, 185, 187, 188, 189, 190, 192, 193, 194, 195, 196, 198, 200, 201, 202]	125	7%

personalization, goal-setting, rewards, and other social support strategies. Furthermore, if we consider the employment of all social support strategies as overall (e.g., social learning, social cooperation, social comparison, social competition, normative influence, social facilitation, social recognition, and other social support strategies), we can notice that the second most effective and commonly employed set of strategies were social support strategies which were mainly used as external motivations to persuade users to engage more in increasing their PA levels and reducing SB.

4.7. Behavior Theories Employed and the Effectiveness of PTs

Evaluating the studies based on the behavior theories they employed shows that 125 studies, approximately three quarters (74%) did not have any theory informing their design of the PTs, as shown in **Table 7** and **Figure 10A**. However, among the studies employing theories, many of the studies also only mentioned the theories without providing details of how they informed the study and design of the PT. **Table 7** and **Figure 10A** show that the Transtheoretical model of change (TTM) was the most frequently

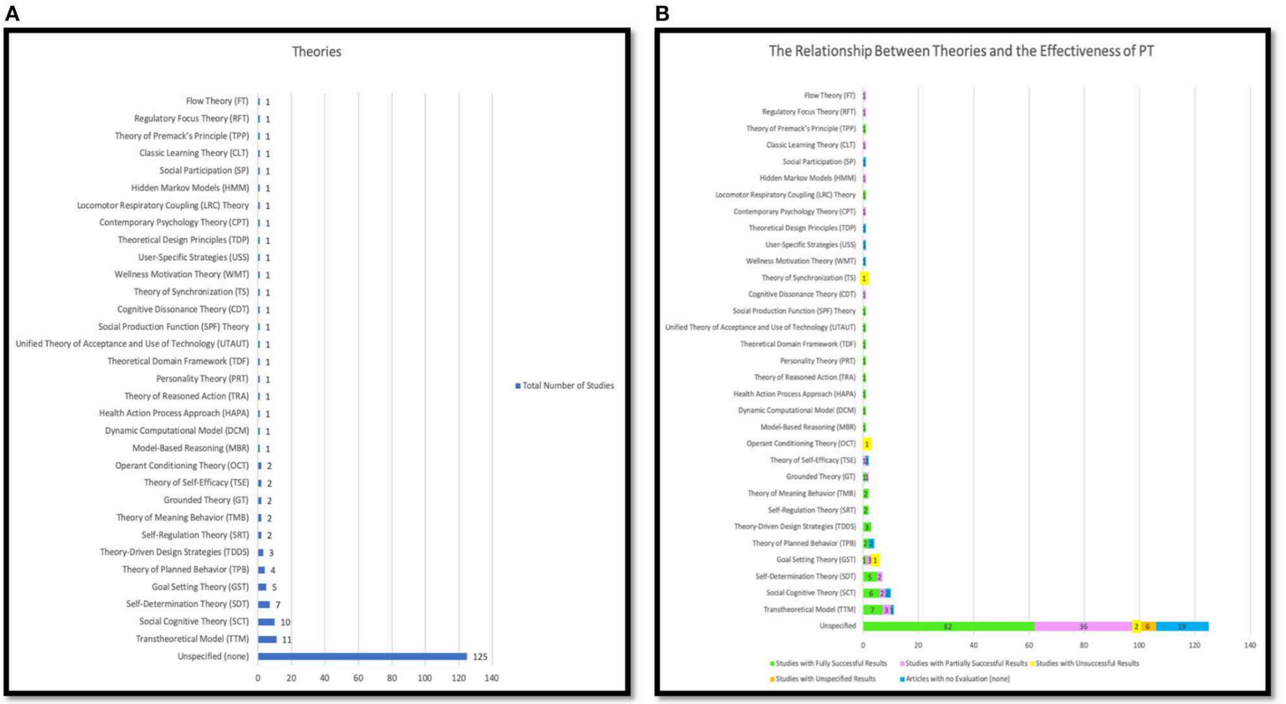


FIGURE 10 | (A) Behavior theories used in persuasive technology design; **(B)** the relationship between behavior theory and the effectiveness of PT.

employed in the studies that were analyzed, with a total of 11 (6%) of studies. Social cognition theory and self-determination theory were second and third with a total of 10 (6%) and 7 (4%) studies respectively. Furthermore, many of the studies used more than one theory, or adapted more than one theory to guide the PT design.

As shown in **Figure 10B**, based on our analysis, a total of 98 (78%) of all the studies employing no theory reported successful outcomes, whether fully or partially successful, while only 2 (2%) reported unsuccessful results. Nineteen of the studies that did not employ any theory conducted no evaluations. With respect to the studies employing theories (45 studies), 39 (86%) reported successful results, whether fully or partially successful, while 2(4%) reported unsuccessful results. Four of the studies that employed theories conducted no evaluations. We could not precisely compare the effectiveness of PTs employing behavior theories and those that did not because of the limited number of studies employing theory. However, we noticed that although limited, the studies employing theory in their design seem to be more effective compared to those that are not based on any theory.

4.8. Targeted Health Behavior Domain

In this study, all the articles selected for review were those that targeted PA and/or SB. **Table 8** and **Figure 11** illustrate how we categorized the health domains in this paper into three groups based on the main objective of each study. One hundred and five (62%) of the studies focused on increasing physical activity (PA) levels, and 47 (28%) studies focused on mitigating sedentary behavior. Eighteen (11%) studies aimed to both increase PA levels and reduce SB. In general, the domains covered in this paper are classified into two main categories, PA and/or SB.

TABLE 8 | Targeted health domains.

Domain	Total	Overall of 170
Physical Activity (PA)	105	62%
Sedentary Behavior (SB)	47	28%
Mixed PA and SB	18	11%

4.9. Targeted Behavioral and/or Psychological Outcomes

Table 9 and **Figure 12** display the behavioral and psychological outcomes targeted by the reviewed articles.The articles targeted 21 diverse outcomes as most of the reviewed studies targeted more than one behavioral and/or psychological outcome. Almost three quarters of the studies 151 (89%) were targeted at actual behavior change, which consists of promoting/encouraging a shift from undesirable behavior and habit [4], promoting physical activity and discouraging SB. We found that 51 (30%) of the studies targeted a change in motivation, 42 (25%) increased the awareness for the users, and 11 (6%) focused on changing the attitude of the individuals. Several of the studies targeted the emotions, loneliness, adherence, intentions, and self-efficacy of the individual, as shown in **Table 9** and **Figure 12**. The category "Unspecified" refers to those studies that did not specify the

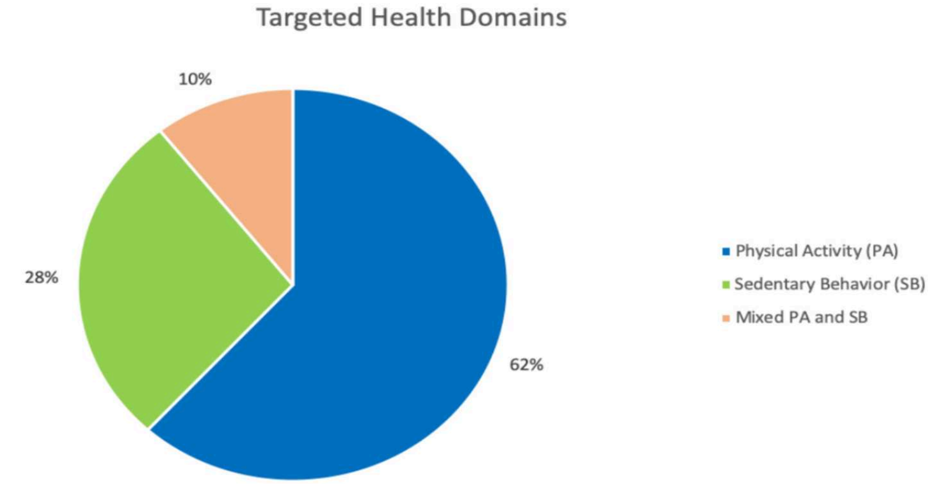


FIGURE 11 | Targeted health domain.

targeted behavioral and/or psychological outcomes. It is also important to note that most of the studies targeted more than one behavioral outcome, which means that many of the studies belonged to more than one category. For example, one study could be targeting the behavior and attitude change in the user.

4.10. Study Methodology Used by Persuasive Technology

Table 10 and **Figure 13A** demonstrate the frequency of the study methodologies used by PTs in the reviewed studies. The quantitative method was the most common methodology employed in the reviewed studies, with a total of 68 (40%). The mixed method was the second most common method used, with a total of 51 (30%) studies. Of the reviewed studies, 28 (16%) studies used a fully qualitative method. The most common quantitative approaches used for data gathering were the use of activity trackers, monitors, and sensors devices, and the use of other systems capable of gathering quantitative data of users' behaviors such as step counters. Moreover, the questionnaire/survey was used as a quantitative method to collect numeric data. The qualitative methods used in the PA and/or SB studies include observations of users' performance, interviews, and focus groups. Although the mixed method (a combination of qualitative and quantitative methodologies) ranked as the second most commonly employed evaluation approach, it is considered as the most comprehensive approach to analyzing the PT design outcomes. Therefore, we recommend researchers to apply the mixed evaluation methodology over a qualitative methodology alone or a quantitative methodology alone.

4.11. Evaluation Methods and Persuasive Technology Effectiveness

As **Table 11** and **Figure 13B** illustrate, out of the 68 studies that employed a quantitative evaluation, 46 (68%) reported fully successful outcome, 18 (27%) partially successful outcomes, 1 (1%) an unsuccessful study, and 3 (4%) were studies that did not specify their outcomes. However, of the studies that used the mix of quantitative and qualitative evaluation methodologies, a total of 28 (55%) were fully successful, 23 (45%) were partially successful, and 2 (4%) were unsuccessful. The studies which implemented just a qualitative methodology have the least effective outcomes, with a total of 14 (50%) completely successful studies, 11 (39%) partially successful, 1 (4%) unsuccessful, and 2 (7%) included studies with unspecified outcomes.

4.12. Study Participants and Sample Size

The sample size varies greatly among the studies reviewed, as the mean number of subjects was 798 with a minimum of one subject and a maximum of 129,010 participants. There are also some studies that did not report the total number of participants, whereas others also had varying sample sizes at different stages of the PT evaluation. **Table 12** and **Figure 14A** show the targeted audience by age demographic, whereas **Table 13A** and **Figure 14B** present the effectiveness of the interventions depending on the targeted audience. We found that most of the studies (94, or 53%) targeted the adults with most of them reporting successful results. This was followed by 21 (12%) studies that targeted young adults and elderly people. However, only 13 studies (8%) targeted children, 8 studies (5%) targeted teenager, and 2 studies (1%) targeted young children. We also found 17 (10%) studies that did not specify their audience. The most targeted populations were adults and young adults, while the least were older people, children, teenagers, and young children.

Young children include kids in the age group 4 to 7, children in the age group 8 to 12, teenagers from 13 to 17 years old and young adults from 18 to around 30 years old. Adults have a wide age range and could start from 31 to 49 years old, whereas the elderly were 50 years old and above.

4.13. Effectiveness of PTs Based Targeted Audience's Age Group

Tables 13A,B and **Figure 14B** demonstrate the effectiveness of employing PT with regards to the targeted audience's age group.

TABLE 9 | Targeted psychological and behavioral outcomes by persuasive technology.

Targeted outcomes	Study	Total number of studies	Average out of % 170 for each
Behavior	[42, 52, 60, 61, 75, 62, 73, 74, 92, 132, 144, 177, 178, 179, 180]; [198] [34, 58, 35, 36, 39, 70, 41, 51, 43, 44, 45, 46, 48, 53, 55, 56, 57, 59, 63, 64, 65, 66, 67, 68, 69, 71, 77, 78, 118, 79, 80, 81, 82, 83, 84, 85, 86, 108, 87, 88, 111, 89, 90, 91, 93, 94, 95, 96, 151, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 109, 110, 112, 113, 195, 114, 115, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 133, 134, 130, 135, 186, 136, 131, 137, 138, 139, 140, 141, 142, 145, 147, 148, 149, 150, 152, 174, 153, 154, 155, 156, 157, 158, 146, 160, 161, 162, 163, 164, 165, 167, 168, 169, 170, 171, 172, 173, 175, 176, 182, 183, 184, 185, 187, 189, 190, 191, 192, 193, 194, 196, 197, 199, 202, 203]	151	89%
Awareness	[35, 39, 70, 42, 43, 45, 47, 49, 69, 80, 81, 82, 87, 94, 95, 99, 112, 113, 115, 123, 138, 139, 140, 141, 144, 145, 158, 161, 169, 170, 172, 175, 184, 187, 191, 192, 193, 200] [41, 67, 202, 203]	42	25%
Motivation	[36, 37, 42, 44, 47, 50, 52, 53, 57, 60, 61, 75, 65, 69, 81, 95, 100, 103, 110, 111, 112, 113, 114, 116, 119, 129, 134, 136, 141, 143, 145, 162, 166, 171, 177, 178, 179, 181, 183, 185, 186, 189, 191, 192, 193, 194]; [195, 197, 198, 199, 200]	51	30%
Self-management	[159, 187]	2	1%
Attitude	[39, 54, 69, 76, 103, 117, 122, 132, 191, 192, 193]	11	6%
Adherence	[74, 182, 188, 196]	4	2%
Intentions	[190]	1	1%
Cognitive	[72, 83]	2	1%
Physical abilities	[72]	1	1%
Feasibility	[154]	1	1%
Acceptance	[38, 154, 182, 188, 192, 193]	6	4%
Confidence	[194]	1	1 %
Emotion	[56]	1	1 %
Self-Esteem	[56]	1	1 %
Thermal comfort	[201]	1	1 %
Loneliness	[56]	1	1 %
Balance	[56]	1	1 %
Engagement	[48]	1	1%
Perspective	[174]	1	1%
Reducing sitting time	[126]	1	1%
Self-efficacy	[36, 57, 194]	3	2%
Unspecified	[40]	1	1%

For adults, we found that 81(86%) of the studies reported successful results; that is, studies with partially successful and those with fully successful results. Specifically, 47 (58%) studies were fully successful, and 34 (42%) studies were partially successful. For young adults, out of 21 studies targeted at them, 13 (61%) showed fully successful outcomes, 3 (14%) displayed partially successful outcomes, just 2 (10%) reported unsuccessful outcomes, and only 1 (5%) represented unspecified outcomes, and 2 (10%) did not provide evaluations. For the elderly, out of 21 studies targeted at them, 7 (33%) reported fully successful results, 10 (48%) showed partially successful results, only 1 (5%) reported unspecified results, and 3 (14%) did not evaluate their studies. For children, out of 13 studies targeted at them, 8 (62%) reported fully successful results, 3 (23%) showed partially successful results, and just 2 (15%) did not evaluate their PTs. For teenagers, out of 8 studies targeted at them, 4 (50%) reported fully successful results, 2 (25%) provided partially successful results, and only 2 (25%) did not conduct any evaluations. Only two studies provided fully successful outcomes for young children. Therefore, the most successful outcomes for implementing the PTs were observed in the studies targeting adults and young adults.

4.14. Targeted Audience by Their Occupation/Status or Health Condition

Another classification of the targeted audience was based on the audience's situation, such as their occupation and health conditions, as we found from the reviewed studies. As **Table 14** and **Figure 15A** show, 97 (57%) studies did not specify their sample population's status. Thirty-three (19%) studies are targeted at office workers, 11 (6%) at students (e.g., primary school students, and high school students), 6 (4%) at university students (e.g., undergraduate students, and graduate students), 4 (2%) at university workers (e.g., university staff and faculty members), and 3 (2%) studies were targeted at people with overweight and obesity conditions. Nurses, researchers, runners, employees, heavy computer users, medical specialists, patients with type 2 diabetes, patients with chronic obstructive pulmonary disease, breast cancer survivors, arthritis patients, patients with autism spectrum disorder, Fitbit users, older cancer survivors, individuals with severe mental health problems, breast cancer patients, and people with multiple sclerosis were the target of one (1%) of the study each. Consequently, approximately a quarter of all studies focused on office workers because these populations are more likely to remain sitting on their seats and working on their desks for long hours. In such a situation, it is possible for them to suffer some lifestyle-related health issues such as cardiovascular disease, cancer, obesity, and diabetes.

4.15. Duration of Evaluation

The duration of the studies varied from 1 day to ∼2 years. In addition, 23 (14%) studies did not report how long they evaluated their persuasive technologies. The results indicate that 46 (27%) studies evaluated the PT from 1 to 3 months, 27 (16%) studies for <1 week, and 17 (10%) studies for <1 month, 14 (8%) studies for <2 weeks, and 11 (6%) studies

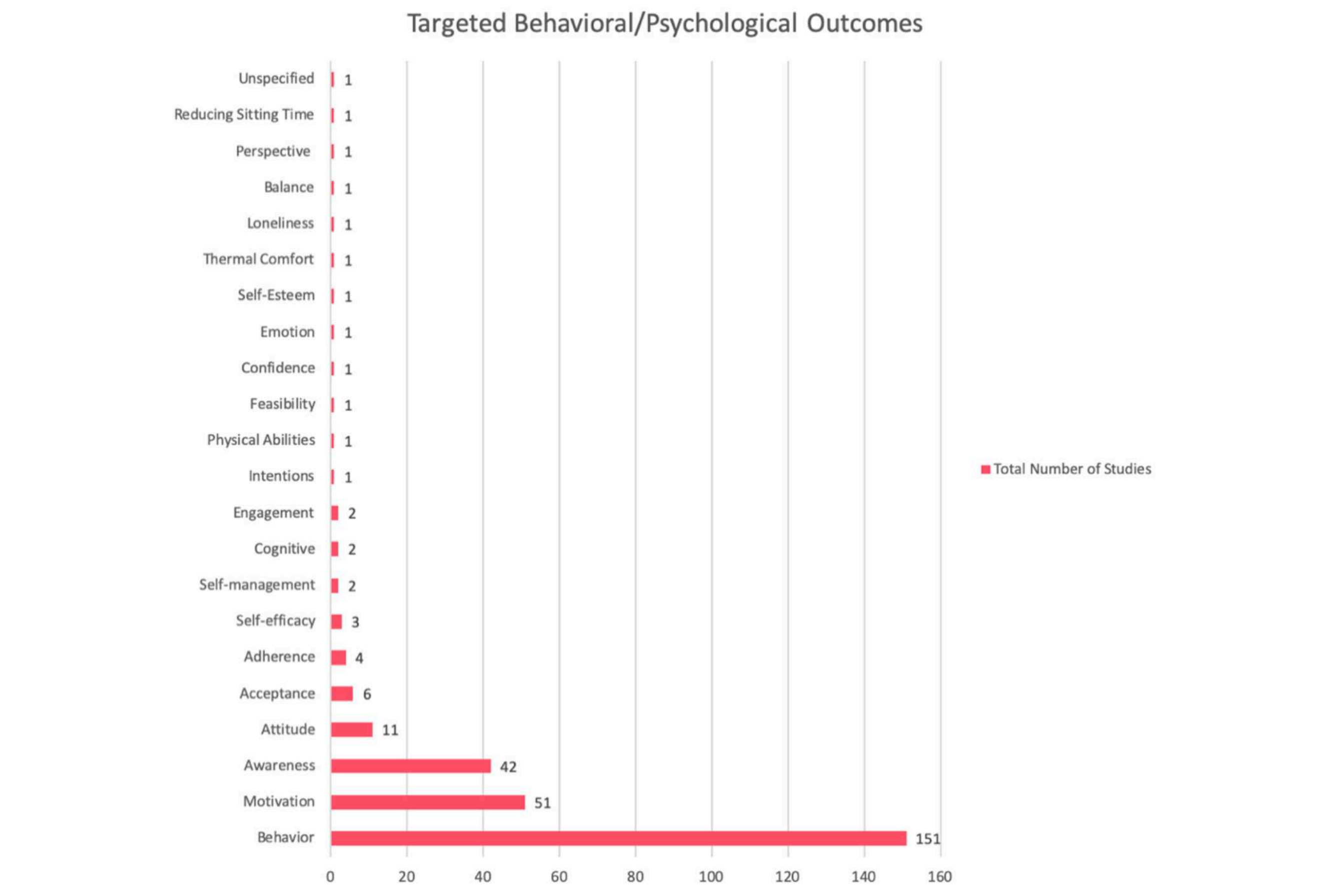


FIGURE 12 | Targeted psychological and behavioral outcomes of persuasive technologies.

TABLE 10 | Evaluation methods and persuasive technology outcomes.

Evaluation method	Number of studies with fully successful results	Number of studies with partially successful results	Number of studies with unsuccessful results	Number of studies with unspecified results	Total	Overall of % 170
Quantitative	46 (68%)	18 (27%)	1 (1%)	3 (4%)	68	40%
Qualitative	14 (50%)	11 (39%)	1 (4%)	2 (7%)	28	16%
Mixed method (Quantitative & Qualitative)	28 (55%)	23 (45%)	2 (4%)	0	51	30%
Number of articles with no evaluation (none)		23			23	14%

for four to six months, and just 2 (1%) studies for <1 year. Only 5 (3%) studies conducted their long-term "longitudinal" evaluations of the effectiveness of PTs for one to a one and a half years, and 2 (1%) for 2 years. The results also reveal that 20 (11%) studies with a longitudinal evaluation have a duration from 4 months to 2 years, whereas 104 (61%) studies conducted their PTs over a duration from <1 week to 3 months. The variation in the duration of evaluating the PTs presents a challenge because it is difficult to establish the long-term effects of the PTs since many studies did not conduct an adequate evaluation and follow-up studies. Consequently, there is still a need to conduct more long-term evaluations of PTs design in the domain of PA and/or SB to examine users' adherence and commitment and establish PTs effectiveness over a long-term for sustained behavior change. **Figure 15B** presents the duration of the evaluation of the reviewed studies.

5. DISCUSSION

The purpose of this study is (1) to evaluate the effectiveness of PT used to promote PA and reduce SB; (2) to summarize and highlight trends in the outcomes and employed technological platforms; and (3) to reveal pitfalls and gaps in the present

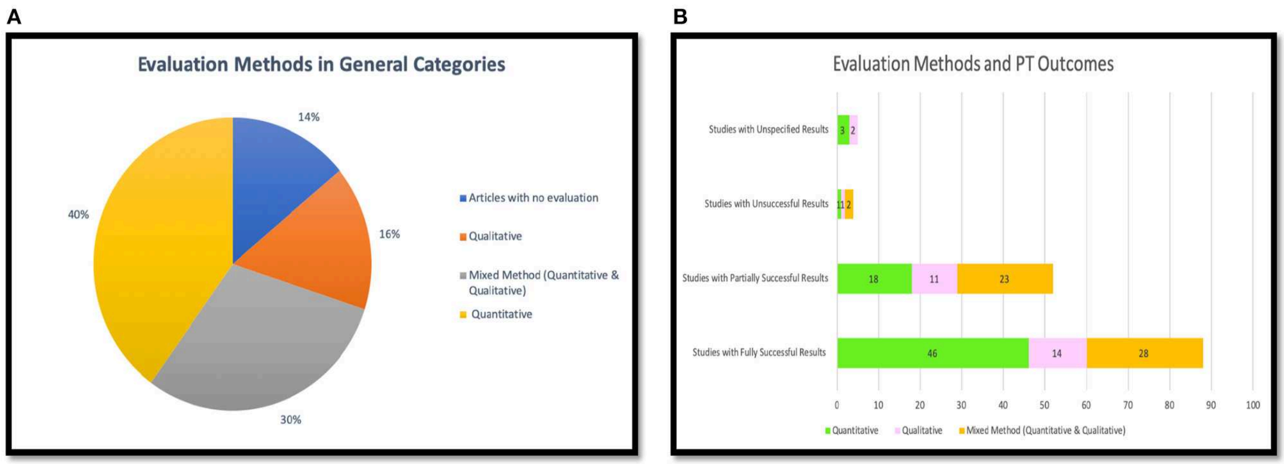


FIGURE 13 | (A) Evaluation methdologies employed by persuasive technology; **(B)** Evaluation methods and persuasive technology effectiveness.

literature that could be leveraged and used to inform the design of PT targeting physical activity and sedentary behavior.

5.1. Overall Effectiveness of PTs for Physical Activity and Sedentary Behavior

Overall, 137 (81%) of the articles that we reviewed in this study reported successful outcomes, whether fully or partially successful, which prove that PTs are effective tools to promote PA and decrease SB. Only 4 (2%) of the reviewed studies had unsuccessful outcomes. There were no common or specific reasons for the failure outcomes of these studies. Each study had a different situation and employed a different method, strategies, and technology that may contribute to unsuccessful outcomes. For example, one study failed in designing an appropriate smartphone virtual boat racing game to motivate people to engage more in Moderate-Intensity Physical Activity (MIPA). This is because the game was not implemented optimally, which caused users to suffer from some repetitive strain injuries and drove them to abandon the app [91]. Other studies implemented different technologies such as Persuasive Art reflection [107], and ExerSync by considering a rhythm of body movements [108]. Therefore, it is difficult to establish the actual reasons for the ineffectiveness of the PTs that reported unsuccessful results. Other reasons may be the target audience, their behavior change stage, and persuasive strategy mismatch, as highlighted [204].

5.2. The Relationship Between Technology Platforms and the Effectiveness of PTs

Mobile and handheld devices were the most dominant technology platforms used, with a total of 61 (36%) studies, followed by games, web and social networks, using of commercially available sensors and other activity trackers, custom-designed sensors and activity trackers, and ambient and public displays, which had a total of 33 (19 %), 32 (19%), 31 (18%), 19 (11%), and 16 (9%) studies respectively (see **Figure 4A**). Therefore, it is very clear that the second most dominant technologies employed in the reviewed studies were sensors and activity trackers and monitors devices, with a total of 50 (29%), either by using commercially available devices or designing new ones. In fact, if we consider the use of the embedded sensors in the smartphones and handheld devices such as GPS, GSM, gyroscope, accelerometer, pedometers, and cameras, we notice that the most important factor to motivate users in doing PA is to give accurate feedback and result of their activities tracked using sensors and activity trackers and monitors. This corroborates our findings, whereby the tracking and self-monitoring strategies ranked first with a total of 153 (90%) studies, of which 121 (79%) reported fully and/or partially successful outcomes, and the reminder strategy ranked second with a total of 72 (42%) studies, of which 32 (44%) had fully successful outcomes, and 27 (38%) had partially successful outcomes. These results suggest that a simple nudge such as a reminder to get some exercise (e.g., take some walk) or about how long they have been sitting down and the need to get up could motivate people to increase their physical activity. This is understandable, considering that in this modern time, people are always busy. So, even when they have the good intention to exercise and also know the consequences of living a sedentary lifestyle, they can easily forget. Therefore, a simple reminder could go a long way, motivating them to action.

As shown in **Figure 4B**, we found that the most successful outcomes for implementing the PTs were observed in the studies using the mobile and handheld devices, games, sensors and activity trackers in general, and websites and social networking sites (SNSs). It seems that these technologies are attractive and promising technologies for delivering interventions because of their ubiquitous nature.

5.3. The Relationship Between Behavior Theory and the Effectiveness of PT

As shown in **Figure 10B**, the findings reveal that almost three quarters 125 (74%) of all the reviewed articles did not use or did not state very clearly the behavior theory they used. Considering that most of the analyzed studies either did not

TABLE 11 | Evaluation methodologies employed by persuasive technology and PT effectiveness.

Evaluation method	Studies with fully successful results	Studies with partially successful results	Studies with unsuccessful results	Studies with unspecified results	Total	Overall of % 170
Quantitative	[37, 38, 45, 47, 50, 56, 57, 61, 75, 62, 68, 73, 79, 89, 92, 93, 94, 95, 109, 114, 117, 118, 124, 126, 128, 130, 136, 131, 141, 144, 147, 148, 149, 151, 155, 156, 158, 146, 163, 166, 169, 173, 176, 196, 197, 203]	[53, 59, 71, 77, 96, 105, 113, 120, 125, 135, 140, 142, 143, 164, 170, 177, 181, 201]	[165]	[98, 138, 152]	68	40%
Qualitative	[49, 50, 58, 60, 67, 80, 110, 121, 153, 154, 174, 175, 188]	[36, 43, 81, 84, 99, 123, 132, 172, 183, 187, 191]	[91]	[55, 97]	28	16%
Mixed method (Quantitative & Qualitative)	[34, 35, 39, 70, 40, 41, 44, 46, 48, 54, 63, 64, 66, 69, 72, 76, 83, 85, 86, 100, 103, 115, 116, 122, 137, 139, 157, 184]	[42, 51, 52, 101, 104, 112, 129, 133, 134, 160, 167, 178, 179, 180, 182, 185, 186, 190, 192, 193, 195, 199, 202]	[107, 108]		51	30%
Articles with no evaluation	[65, 74, 78, 82, 87, 88, 111, 90, 102, 106, 119, 127, 145, 150, 159, 161, 162, 168, 171, 189, 194, 198, 200]				23	14%

Conference Name:ACM Woodstock conference.

TABLE 12 | Targeted audience by age demographic.

Audience category	Study	Total number of studies	Average out of % 170 for each
Young children	[66, 83]	2	1%
Children	[36, 61, 75, 64, 67, 72, 101, 104, 122, 141, 193, 198, 200]	13	8%
Teenagers	[41, 51, 50, 60, 68, 82, 192, 198]	8	5%
Young adults	[37, 44, 45, 46, 55, 76, 82, 85, 86, 108, 102, 107, 134, 136, 148, 153, 155, 175, 181, 192, 203]	21	12%
Adults	[34, 58, 35, 36, 38, 39, 70, 42, 43, 47, 49, 53, 54, 57, 59, 62, 63, 65, 69, 71, 73, 77, 79, 80, 84, 87, 89, 91, 92, 93, 94, 95, 96, 151, 97, 98, 99, 103, 105, 109, 111, 113, 195, 114, 115, 116, 117, 120, 121, 123, 125, 126, 128, 129, 132, 130, 135, 186, 131, 137, 138, 139, 140, 142, 143, 144, 147, 149, 152, 174, 154, 158, 146, 161, 163, 164, 165, 167, 168, 170, 172, 173, 176, 178, 179, 184, 185, 189, 190, 191, 192, 199, 201, 202]	94	55%
Elderly	[52, 166, 177, 180] [40, 48, 56, 78, 81, 100, 106, 112, 119, 133, 157, 160, 182, 187, 188, 192, 197]	21	12%
Unspecified	[74, 88, 90, 110, 118, 124, 127, 145, 150, 156, 159, 162, 169, 171, 183, 194, 196]	17	10%

specify the theories used to inform their design or did not use any theory, it is hard to draw conclusions on the relationship between employing behavior theory and the effectiveness of PTs. However, based on what we have, a total of 98 (78%) of all the studies employing no theory reported successful outcomes, whether fully or partially successful, while only 2 (2%) reported unsuccessful results. Nineteen of the studies that did not employ any theory conducted no evaluations. With respect to the studies employing theories (45 studies), 39(86%) reported successful results, whether fully or partially successful, while 2(4%) reported unsuccessful results. Four of the studies that employed theories conducted no evaluations. Based on this, it seems that the use of behavioral theories to inform PTs design increases the effectiveness of PTs with respect to achieving the intended objective of promoting PA or reducing SB.

5.4. Targeted Outcomes of Persuasive Technology

Most of the studies 151 (89%) targeted actual behavior change in the participants by increasing their level of physical activity, such as increasing step counts. User motivation was the second targeted outcome with a total of 51 (30%) studies, followed by articles that aimed at creating awareness and attitude change in users with totals of 42 (25%) and 11 (6%) studies, respectively. Nevertheless, there are some studies that targeted more than one behavioral or psychological outcomes.

5.5. The Relationship Between Persuasive Strategies and the Effectiveness of PTs

In the present review, various persuasive strategies were identified that were used to achieve positive behavior change.

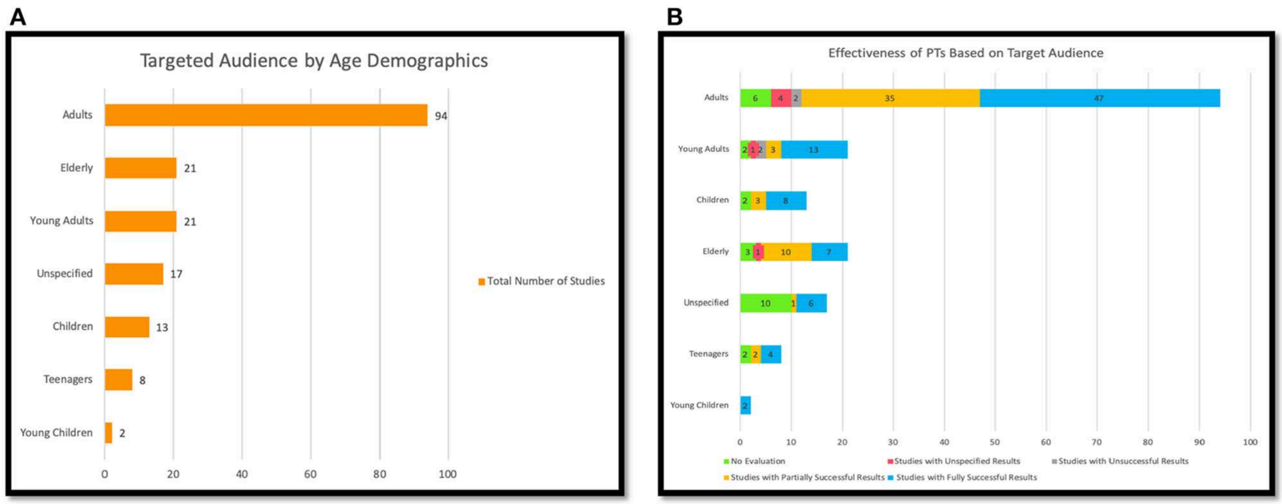


FIGURE 14 | (A) Targeted audience by age demographics; **(B)** Effectiveness and evaluation outcomes of PTs based on target audience.

With respect to the studies employing persuasive strategies, and reported successful results, whether fully or partially, we found the most common strategies employed were tracking and self-monitoring with a total of 153 (90%) studies, of which 121 (79%) were successful studies. The implementation of such strategies was achieved by the use of diverse activity tracking and monitoring devices and sensors such as accelerometers, pedometers, heart rate monitoring devices and embedded sensors in smartphones, and by providing the user with his/her activity performance (e.g., step counts, heart rate, speed, summary progress) on the screen of the mobile phone devices using various display formats including visualization. PTs that used social support strategies (e.g., social comparison, social cooperation, social competition, normative influence, social facilitation, social learning, social recognition, and other social support strategies) were also effective in promoting physical activity with a total of 131 (77%) studies, of which 104 (79%) were studies of successful outcomes involving fully and partially. Overall, other strategies that were effective in addressing PA and SB includes: starting from the most effective to the least and out of the total studies that employed each persuasive strategy: reminders, personalization, goal setting, rewards, simulation, praise, reduction, suggestion, tailoring, tunneling, and expertise with a total of 59 (82%), 58 (90%), 44 (83%), 41 (76%), 31 (74%), 30 (79%), 26 (81%), 26 (87%), 26 (90%), 23 (92%), 20 (80%) and 11(85%) successful studies (whether fully or partially successful), respectively. These strategies were useful in encouraging users to make the appropriate changes in their behaviors and to be more aware and motivated.

It is also necessary to highlight the fact that most of the PT systems employed more than one strategy to achieve the targeted behavioral outcome. Also, the operationalization and implementation of these strategies varied from one application to another and may contribute to the effectiveness of the strategies. For example, some studies used a social support strategy as well as tracking, whereas others used the goal setting and reminder as different motivational strategies. In addition, the self-monitoring strategy came in various implementations, including graphical display, audio, textual, and visual feedback, ambient displays mirror, ambient sculpture display, and light displays.

Furthermore, the key implication from our findings is that there are considerable discrepancies in naming and implementing the persuasive strategies in the PT systems reviewed. Some PTs also implemented strategies that are not captured in the existing PSD framework. This makes it difficult to easily extract, identify, and name the strategies employed in PT. This makes the identification of such strategies to be based merely on the researchers' perspectives of PT. Although, there were diverse accomplishments in the research field in designing models that identify, classify, and name various persuasive strategies and their functionalities [205, 3]. Existing frameworks appear not to be comprehensive enough to capture all possible strategies in this considering the fast advancement of technology and opportunities that it creates to use various technology-enabled strategies that were probably not possible when existing models were developed. Therefore, we suggest that more work is needed in the area of developing a comprehensive PT design framework that captures all possible design strategies and various ways each can be operationalized in PT designs to achieve the desired behavioral outcome.

These findings agree with Orji and Moffatt [4]. As aforementioned, the persuasive system design (PSD) model seems not comprehensive enough to identify and classify all the strategies. As a result, we identify more strategies that were not included in the PSD.

5.6. The Relationship Between Targeted Audience and the Effectiveness of PT

Many PTs have been employed to persuade different age groups of users to change or adopt a desirable lifestyle with regard to

TABLE 13 | (A) Targeted audience by age group and persuasive technologies effectiveness; **(B)** Effectiveness of persuasive technologies based on targeted audience.

(A)

Targeted audience by age group	Age group (years old)	Studies with fully successful results	Studies with partially successful results	Studies with unsuccessful results	Studies with unspecified results	Articles with no evaluation	Total number of studies	Average out of 170 for each
Young children	4 to 7	[66, 83]					2	1%
Children	8 to 12	[61, 75, 64, 72, 101, 122, 141] [67]	[36, 104, 193]			[198, 200]	13	8%
Teenagers	13 to 17	[41, 50, 60, 68]	[51, 192]			[82, 198]	8	5%
Young adults	18 to 30	[37, 44, 45, 46, 76, 85, 86, 136, 148, 153, 155, 175, 203]	[134, 181, 192]	[107, 108]	[55]	[82, 102],	21	12%
Adults	31 to 49	[34, 58, 35, 38, 39, 70, 47, 49, 54, 57, 62, 63, 69, 73, 79, 80, 89, 92, 93, 94, 95, 103, 109, 114, 115, 116, 117, 121, 126, 128, 130, 131, 137, 139, 144, 147, 149, 151, 154, 158, 146, 163, 173, 174, 176, 184, 191]	[36, 42, 43, 53, 59, 71, 77, 84, 96, 99, 105, 113, 195, 120, 123, 125, 129, 132, 135, 186, 140, 142, 143, 164, 167, 170, 172, 178, 179, 185, 190, 192, 199, 201, 202]	[91, 165]	[97, 98, 138, 152]	[65, 87, 111, 161, 168, 189]	94	55%
Elderly	50 and above	[48, 56, 100, 157, 166, 188, 197]	[52, 81, 112, 133, 160, 177, 180, 182, 187, 192]		[40]	[78, 106, 119]	21	12%
Unspecified	Not specified	[110, 118, 124, 156, 169, 196]	[183]			[74, 88, 90, 127, 145, 150, 159, 162, 171, 194]	17	10%

(B)

Targeted audience by age group	Number of studies with fully successful results	Number of studies with partially successful results	Number of studies with unsuccessful results	Number of studies with unspecified results	Number of articles with no study (none)
Young children (4 to 7)	2	0	0	0	0
Children (8 to 12)	8	3	0	0	2
Teenagers (13 to 17)	4	1	0	0	2
Young adults (18 to 30)	13	3	2	1	2
Adults (31 to 49)	47	34	2	4	6
Elderly (50 and above)	7	6	0	1	3
Unspecified	6	0	0	0	10

TABLE 14 | Audience occupation /status/ health conditions.

Audience occupation or health condition	Study	Total number of studies	Average out of % 170 for each
School students	[50, 72, 75, 94, 101, 104, 122, 141, 181, 193, 200]	11	6%
University students (Undergraduate Students, Graduate students)	[86, 123, 134, 179, 190, 202]	6	4%
Office workers	[35, 43, 49, 71, 73, 79, 89, 93, 95, 98, 99, 109, 111, 113, 195, 126, 133, 130, 142, 147, 149, 151, 152, 174, 154, 157, 146, 161, 163, 164, 165, 167, 201]	33	19%
Nurses	[63]	1	1%
University workers (Information workers as University staff members, student council) and other workers	[96, 105, 178, 190]	4	2%
Patients with type 2 diabetes	[140]	1	1%
Medical specialists	[138]	1	1%
Heavy computer users	[116]	1	1%
Researchers	[94]	1	1%
Overweight & obese individuals	[59, 114, 166]	3	2%
Working employee	[77, 153]	2	1%
Patients with chronic obstructive pulmonary disease (CODP)	[182]	1	1%
Arthritis patients	[187]	1	1%
Breast cancer survivors	[191]	1	1%
Patients with autism spectrum disorder (ASD)	[189]	1	1%
Fitbit users	[186]	1	1%
Athletes	[143]	1	1%
Older cancer survivors (OCS)	[100]	1	1%
Individuals with severe mental health problems	[185]	1	1%
Runners people	[132]	1	1%
Breast cancer patients	[168]	1	1%
People with multiple sclerosis	[197]	1	1%
Diverse occupations (e.g., Administrator, Human resources specialist, Economist, Engineer, Educator, & real estate agent)	[129, 179]	2	1%
Unspecified	[34, 58, 36, 37, 38, 39, 70, 40, 41, 51, 42, 44, 45, 46, 47, 48, 52, 53, 54, 55, 56, 57, 60, 61, 62, 64, 65, 66, 67, 68, 69, 74, 76, 78, 118, 80, 81, 82, 83, 84, 85, 87, 88, 90, 91, 92, 97, 102, 103, 106, 107, 108, 110, 112, 115, 117, 119, 120, 121, 124, 125, 127, 128, 135, 136, 131, 137, 139, 144, 145, 148, 150, 155, 156, 158, 159, 160, 162, 169, 170, 171, 172, 173, 175, 176, 177, 180, 183, 184, 188, 192, 194, 196, 198, 199, 203]	97	57%

PA. As displayed in **Figure 14B**, **Tables 12, 13A**, the reviewed studies showed that PT targeted at adults recorded the highest success rate, with 82 (87%) successful outcomes, of which 47 (58%) were fully successful outcomes and 34 (42%) were partially successful outcomes. Out of the total studies that targeted each age demographic, the second and third placed were elderly, and young adults, with a total of 17 (81%) and 16 (76%) of successful results studies, respectively. The fourth rank was children with 11 (85%) successful results. The studies that did not specify their target audience ranked 5th with 7 (41%) successful outcomes. The sixth and seventh placed were teenagers and young children with a total of 6 (75%) and 2 (100%) successful outcomes, respectively. As previously mentioned, the present study demonstrates that PT was most effective among adults when targeting PA and SB. However, it is important to note that the majority of the studies evaluated were targeted at the adult population; hence, comparing success rates across populations may not make much sense. A possible reason while most studies targeted adults reported successful results is that adults are in their active stage of life and at this stage, people tend to be more active naturally compare to the elderly group. Again, in comparison to children, adults tend to be more conscious about their life because they have the cognitive ability to understand the consequences of a sedentary lifestyle.

Furthermore, more than half of the reviewed articles did not specify their targeted audience occupation/status or health conditions, totaling 97 (57%) articles. However, 33 (19%) of the total articles were targeted at office workers. We believe that this is due to the nature of their jobs, which often lead to prolonged sitting (e.g., for hours) without taking frequent breaks to do some

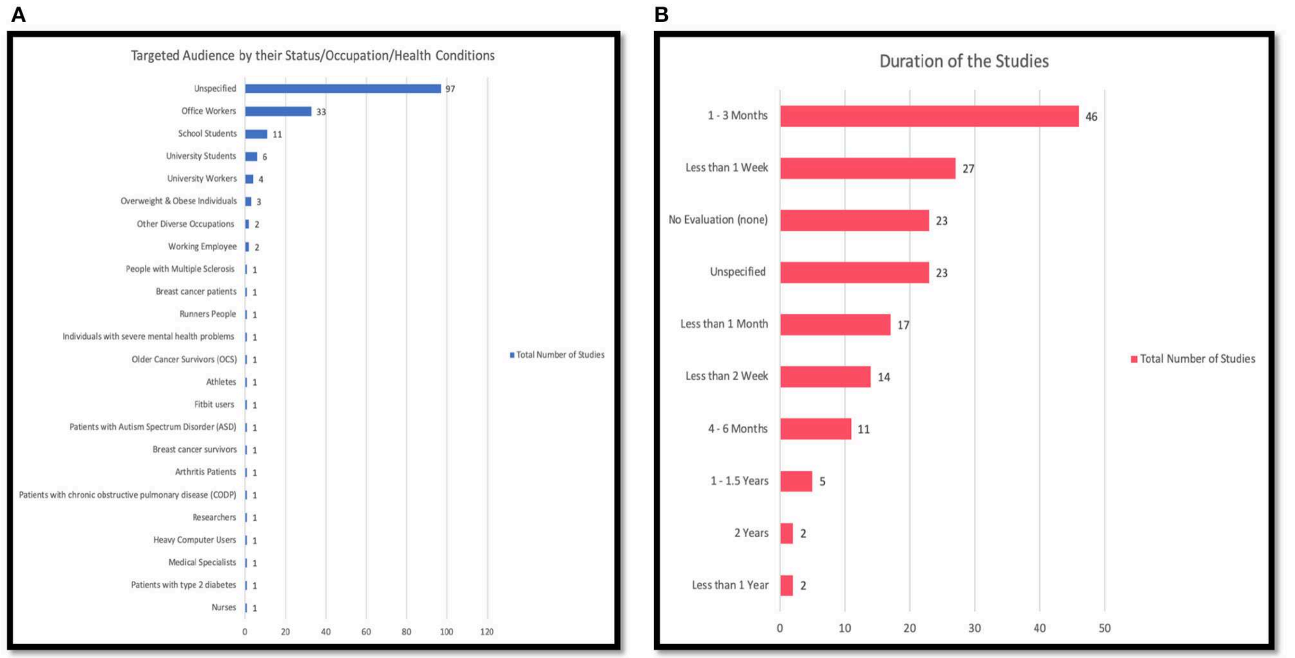


FIGURE 15 | (A) Audience categorization based on their occupation/health conditions; **(B)** Duration of studies' evaluation.

PA, such as stretching and walking. However, these articles did not specify their target audience's health situation beyond their occupation. One reason for this could be because conducting users' studies for the evaluation purposes of PT systems in adults and the general population without stating any conditions or restrictions is easier and more time-saving than conducting a study with a specific sample of users that have restricted criteria or specific health issues.

5.7. General Recommendations for Future Research

The review identified a number of limitations and gaps in the existing works in the area of PT for PA and SB. We offer suggestions for advancing research in this area:

1. **Standard Approach for Evaluating Persuasive Technology**: There is a need for a standard approach for evaluating the effectiveness of PTs, in order to provide standard and reliable data that can be used to inform future PT designs. Most of the studies reviewed presented subjective data with no standard approach by which to measure whether or not the technologies were effective, and to what extent they were effective.
2. **Using Behavior Theories to Inform Persuasive Technology Design**: Although our analysis could not successfully compare the effectiveness of PTs employing behavior theories and those that did not, due to the limited number of studies employing theory. Our result shows that although limited, PTs employing theory in their design tend to be more effective than those based not on any theory, although marginal. This supports previous research suggesting that PTs based on theory are more effective than those based on intuition [206, 207, 208]. A possible reason why most PT designers do not employ theories is probably because most designers lack the necessary background to appropriately interpret behavior theories and translate them into actionable and practical PT design components [208]. Hence, PT designers can collaborate with people that have an adequate background such as behavioral scientists to achieve this. Therefore, *we recommend that PT designer employ behavior change theories in their design and clearly state how the theoretical components were translated into the design components in the PTs.*
3. **Effectiveness of Persuasive Technologies Employing Multiple Strategies vs. Those Based on a Single Strategy**: There is also a need to establish the effectiveness of PTs employing a single persuasive strategy in comparison to those employing multiple strategies. Although, employing multiple strategies has been the convention in the area with the hope that the more the better. However, this may not be the case. As shown by Orji et al. [209], PTs employing a single strategy can be effective. Nevertheless, it is unknown whether employing multiple strategies would result to more effective PTs; that is if the strategies have an additive effects. We also acknowledge that employing multiple strategies may lead to a cognitive overload on the part of the users. Hence, *we recommend that future research should focus on establishing the effectiveness of PTs employing a single strategy in comparison to those employing multiple strategies and also how this may vary depending on how the strategies are implemented.*
4. **Effectiveness of Persuasive Strategies Across Contexts**: Although the review focused on studies in the area of PA and SB, we also noticed a variation in the choice of strategies which

are majorly and randomly chosen due to the lack of clear guideline on which strategy works under various contexts. Hence, *we recommend that the effectiveness of the strategies be evaluated across domains and technologies to establish domain or technology-dependent factors that may impact effectiveness*. That is, does the effectiveness of the strategies depend on the technology platform and/or the domain of application or they generalized? Research in this area would identify the strengths and weaknesses of each strategy based on many factors, including the sample demographics, their health conditions, and the target behavior. This is essential for advancing the field and contributing to the design of future PT.

5. **Mix-Method Approach to Persuasive Technology Evaluation**: Researchers should employ mix methods approach to uncover the full effects of their PTs. Most existing studies employed the quantitative approach, and this is good as it allows for tracking of the actual PA behavior; however, it gives no insight into the process through which PTs motivated users and inspire the observed behavior change. Qualitative methods such as interviews, on the other hand, would allow users to express their feeling and the motives behind their actions. This would give insight into the reasons behind their actions, which would, in turn, shed more light on the mechanism through which PTs promotes behavior change. Hence, *we recommend that designers should employ a combination of quantitative and qualitative approaches (mixed methods) when evaluating the effectiveness of their PTs*.
6. **Longitudinal Evaluation of Persuasive Technology Effectiveness:** More than half of the reviewed studies 104 (61%) conducted their assessment in duration between <1 week and 1-to-3 months, whereas only 20 (12%) of studies conducted longitudinal evaluations between 4 months and 2 years. Therefore, there is a need to conduct more long-term evaluations to establish the effectiveness and users' adherence to PTs over the long-term for a sustained behavior change in the area of PA and SB domains.
7. **Accessible Cross-platform Persuasive Technologies**: A good number of evaluated PTs are multi-platforms PT intervention. They are implemented to run across multiple technology platforms such as a combination of smartphones, activity trackers devices, cameras, and height-adjustable workstations. We cannot state categorically that it contributes to the effectiveness of such interventions, however, it appears to be a good practice only considering that implementing cross-platform PTs increases the accessibility of such PTs, the reach, and makes them always available for users owning multiple technologies. Therefore, *we recommend that PT designers consider designing a cross-platform application to increase their reach and accessibility*.
8. **Comprehensive Persuasive Technology Design Framework**: Existing PT design models and frameworks are not comprehensive to guide the analysis of current PTs. We identified some strategies that are not captured in the popular PSD model. This is possibly due to advancements in technology evolution, which have made many strategies that would not have been imagined a decade ago possible. Therefore, *we suggest that more work is needed in the area of developing a comprehensive PT design framework that identified not only the strategies but also various possible implementation, domain, user group, technology, and other contextual factors that may affect their effectiveness*. This will hence, facilitate tailoring of PTs based on may contextual factors and user type. The PSD model was useful in organizing the strategies, but it was not enough to include all the resulted strategies. Furthermore, we sometimes faced some confusion when using the PSD model strategies because of the similarity between some of its strategies as well as the method of implementing such strategies in the design of PT based on a designer's own intuition. For example, the growth levels and the happy facial expression of the fish in the "Fish in Steps" system can be considered feedback and rewards strategies, whereas a sad or angry facial expression can also be classified as punishments (or negative reinforcements as they are known), reminders, and feedback strategies [39].

   Again, in most cases, we had to study the functionality for most of the strategies in-depth, which many did not specify clearly, requiring extra time and effort to identify them from the articles. They also had different names and classifications, which made it even more difficult to identify and code them into the PSD model.
9. **Unified Standard for Target Audience Categorization**: The classifications of the demographics by their age groups are sometimes unclear. For example, the age group of adults was varied in the reviewed articles, and this is the same with other age groups such as teenagers and children. This may cause considerable confusion when classifying the targeted audience by their age group. Therefore, *we suggest a unified standard for age group categorization*.
10. **Publication Biases**: It is important to consider publication bias and how it may have affected the present review. This means that papers with positive or significant results are more likely to be submitted and published compared to those with negative findings. However, future research may benefit from research that has reported negative findings/complications with the use of PT. Such information may be useful in directing the design of future PT.
11. **Diversification of the Target Audience of Persuasive Technology**: Most of the reviewed studies were targeted at adults, therefore it is necessary to develop more PT systems that target different populations, such as children, teenagers, and the elderly.
12. **Clarity of the Persuasive Technology Design Objectives**: It is clear that there is confusion regarding the PA and SB domains. People may misunderstand the difference between these domains because they might consider that the most common purpose of designing the PT in such fields is very often the same goal when aiming to reduce the time the user spends sedentarily and to increase his/her PA levels. However, it is important for researchers to distinguish between the terminologies of PA and SB. This is because each domain may require PT designers to employ different persuasive strategies or implement the strategies differently to achieve the desired objectives based on the

TABLE 15 | Check list for future design and research of PT for PA and/or SB.

Tailoring PTs targeted population	Researchers have to consciously consider the targeted population by their age demographics, health conditions, jobs, and their general status in designing PT and employing appropriate and suitable persuasive strategies. PTs that are tailored and reflect the target audience realities tend to be more effective than the generic ones that employ the one-size-fits-all approach.
Design approach	To effectively tailor and consider the target audience in PT designs, PT designers should employ the iterative user-centered design approaches which involve studying and engaging the target audience from the onset of the design to the final deployment and evaluation.
Privacy	• It is essential to provide users with their performance feedback, notifications, and progress updates without intruding on their autonomy or privacy. • Users need to have control over which and how data will be tracked and what they will be used for.
Duration of evaluation	Researchers need to conduct longitudinal evaluations for their PT design to assess the effectiveness and users' commitment and adherence in continuing to use the PT over the long term in the area of PA and SB.
PT platforms	Although most existing studies employed multiple technology platforms (e.g., a combination of a wearable activity tracker and mobile phone) in PT design to persuade users to perform more PA and reduce SB, this may be burdensome on the user and discourage long-term use as they may not be seamlessly integrated into user's daily. We suggest that simple PTs based on a single platform that can easily integrate into user's daily lives should be preferred over complex ones that requires combining and carrying many gadgets.
Evaluation approach	Designers should prefer mix-method evaluation that combines both quantitative and qualitative approaches over a single method. This tend to provide a comprehensive evaluation of the PTs, uncovering not only what works but why and how they work
Behavior theories	Studies employing theory in their design tend to be more effective than those based not on any theory. Therefore, we recommend that PT designers employ behavior change theories in their design and clearly state how the theoretical components were translated into the design components in the PTs.
Others	• Researchers need to state specifically the main purpose of their PT design, whether aiming to increase PA alone, reduce SB alone or both. • Researchers need to state clearly the persuasive strategies they employed and how such strategies are implemented in their PT design (e.g., a self-monitoring strategy that was employed as graphical/visual feedback on a smartphone screen). • Researchers need to consider employing one or a set of PT design models and frameworks to guide the analysis of PTs and the persuasive strategies employed.

targeted domain—PA, SB, or both. For instance, PT aimed at motivating users to achieve the Moderate Intensity Physical Activity (MIPA) level (e.g., 30 min of moderate intensity physical activity (MIPA) daily or 150 min of MIPA weekly) may be different from that aimed at motivating users to perform periodic movements (e.g., standing, stretching, walking) every 30 min or every 1 h to avoid a sedentary lifestyle.

Table 15 displays a list of some essential for PT researchers and designers. This alongside the 12 recommendations above could be used to inform future design and analysis of PT for PA and SB.

5.8. Notes for Future Design of Persuasive Technology for PA and SB

An important point to note is that many of the reviewed studies implemented their PT in more than one technology platforms, such as a combination of smartphones, wearable activity trackers' devices, smartwatch, and sensory chairs, therefore each of these can be considered a multi-platform intervention to achieve the main objective of a study to increase PA levels and reduce SB. This seems to be common considering that users tend to own multiple gadgets these days and to ensure that the PT is always available, they may need to be cross-platform, e.g., integrated with both smartwatch and mobile phone. In that way, it presents multiple opportunities to persuade and motivate users. More importantly, it can be used by users owning various technology, technology-independent. However, this means that the overall cost of implementing PT would increase. Users often do not want to be limited by the technology platform. Hence PT designer, especially those targeting PA and SB, should be aware of this.

Another essential point to consider is that most of the PT employed two or more persuasive strategies (e.g., tunneling, self-monitoring, rewards, reminders, expertise, and social comparison) to persuade users to be physically active and to make them more aware of the side effects of being sedentary. This makes it impossible to know which of the employed strategies resulted in the observed behavior change.

Again, it is also essential that PT designers explicitly state the main objective and purpose of their design, whether targeting in increasing PA alone or decreasing SB alone or both. Most times, this is not clear and a reader would have to deduce from the working of the system, study design, and measured evaluation outcome. This makes analyzing existing studies difficult to achieve.

It is important to state that there is a tiny difference between encouragement and persuasion on one side and coercion and deception on another ([210, 211]). Therefore, it is essential to consider this variation in general when designing PTs, and for health and wellness in particular such as PA and SB domains. According to Vlieghe and De Troyer ([211]), there are some ethical considerations of persuasion that need to be considered:

- The app needs to be tailored to the users' needs and deliver feedback, notifications, progress updates, and cues, which if not carefully implemented, may be considered as surveillance. There is a need to balance between the collection of data and the intruding on the autonomy and privacy of the user.
- There is also a need to design PTs that permit the user to control how the data is tracked, and what it is used for. This

is important for all PTs but more important for PTs that track some health data and health-related behavior data.
- The technologies have to be designed in a way that the persuasive design do not lower the users' autonomy.

6. CONCLUSION AND FUTURE WORK

The paper provides a detailed systematic review of 170 paper to establish the effectiveness PTs for promoting health and wellness in the domains of Physical Activity and Sedentary Behavior. Our findings show that almost three quarters [137 studies (80%)] of the total reviewed studies (170 studies) reported successful outcomes, whether fully or partially successful, which means that PTs are effective at promoting (PA) and discouraging (SB). Thus, the findings demonstrate that the use of PT has the potential to promote desirable behavior change among the users when combined with the proper persuasive strategy. Furthermore, the study summarizes and highlights trends in the outcomes including system design, research methods, persuasive strategies and implementations, behavioral theories, and employed technological platforms. The most frequently targeted populations are adults and young adults, while the least are older people, children, teenagers, and young children. The outcomes of this work illustrate that the most two effective and commonly employed technology platforms in the field of PA and/or SB are mobile and handheld devices, and activity trackers and sensors (whether commercially available or custom-designed by researchers).

Furthermore, this study shows that the most effective and frequently implemented persuasive strategies in PT design for promoting PA and/or reducing SB are tracking/self-monitoring, reminders, personalization, goal setting, rewards, and the set of social support strategies, in decreasing order. Our results show that, although limited, the studies employing behavioral theories in their design tend to be more effective and promising than those not based on any theory. In addition, the research shows that applying the mixed evaluation method (a combination of quantitative and qualitative approaches) is more useful to uncover the full effect of PTs. Finally, we identified the pitfalls and gaps in the present literature that could be leveraged and used to inform the design of a PT that targets PA. Accordingly, we provide a list of general limitations and recommendations to advance and improve future research.

Future works may need to evaluate studies done in the field of PTs in promoting PA and SB according to the different targeted populations by age demographics (e.g., older people, teenagers, children). Future works should also conduct more long-term evaluations to establish the effectiveness of and users' adherence to the PT over the long term in the area of PA and SB. Additionally, we suggest analyzing PTs based on each of technology platforms used in their design. Finally, we also recommend evaluating users' reviews/feedback for the existing PTs (e.g., applications, systems, or devices) to advance the future design of PTs for PA and SB.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

NA conducted the paper search, the thematic analysis, and wrote the first version of the manuscript. FA and RO contributed to reviewing and refining the manuscript. RO and SS supervised the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00007/full#supplementary-material

REFERENCES

1. Tremblay, M. S., Aubert, S., Barnes, J. D., Saunders, T. J., Carson, V., Latimer-Cheung, A. E., et al. (2017). Sedentary behavior research network (SBRN)–terminology consensus project process and outcome. *Int. J. Behav. Nutri. Phys. Act.* 14, 1–17. doi: 10.1186/s12966-017-0525-8
2. Dunstan, D. W., Thorp, A. A., and Healy, G. N. (2011). Prolonged sitting: Is it a distinct coronary heart disease risk factor? *Curr. Opin. Cardiol.* 26, 412–419. doi: 10.1097/HCO.0b013e3283496605
3. Oinas-Kukkonen, H., and Harjumaa, M. (2009). Persuasive systems design: key issues, process model, and system features. *Commun. Assoc. Inform. Syst.* 24:96. doi: 10.17705/1CAIS.02428
4. Orji, R., and Moffatt, K. (2018). Persuasive technology for health and wellness: state-of-the-art and emerging trends. *Health Informatics J.* 24, 66–91. doi: 10.1177/1460458216650979
5. Fogg, B. J. (1998). *Persuasive Computers: Perspectives and Research Directions*. Retrieved from: www.captology.org
6. Orji, R., and Mandryk, R. L. (2014). Developing culturally relevant design guidelines for encouraging healthy eating behavior. *Int. J. Hum. Comput. Stud.* 72, 207–223. doi: 10.1016/j.ijhcs.2013.08.012
7. Wang, Y., Wu, L., Lange, J. P., Fadhil, A., and Reiterer, H. (2018). Persuasive technology in reducing prolonged sedentary behavior at work: a systematic review. *Smart Health* 7–8, 19–30. doi: 10.1016/j.smhl.2018.05.002
8. Gardner, B., Smith, L., Lorencatto, F., Hamer, M., and Biddle, S. J. H. (2015). How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. *Health Psychol. Rev.* 10, 89–112. doi: 10.1080/17437199.2015.1082146
9. Healy, G., Lawler, S., Thorp, A., Neuhaus, M., Robson, E., Owen, N., et al. (2012). *Reducing Prolonged Sitting in the Workplace an Evidence Review: Full Report*. Retrieved from: www.vichealth.vic.gov.au/workplace
10. Shrestha, N., Ijaz, S., Kukkonen-Harjula, K., Kumar, S., and Nwankwo, C. P. (2015). Workplace interventions for reducing sitting at work. *Cochrane Database Syst. Rev.* 1:CD010912. doi: 10.1002/14651858.CD010912.pub2
11. Chu, H. Y. S., Ng, H. X., Tan, C. S., Win, A. M., Koh, D., et al. (2016). A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. *Obes. Rev.* 17, 467–481. doi: 10.1111/obr.12388
12. Dunn, E. E., Gainforth, H. L., and Robertson-Wilson, J. E. (2018). Behavior change techniques in mobile applications for sedentary behavior. *Digit. Health* 4:205520761878579. doi: 10.1177/2055207618785798

13. McCallum, C., Rooksby, J., and Cindy, M. G. (2018). Evaluating the impact of physical activity apps and wearables: interdisciplinary review. *JMIR Mhealth Uhealth.* 6:e58. doi: 10.2196/preprints.9054
14. Rao, A. K. (2019). Wearable sensor technology to measure physical activity (PA) in the elderly. *Curr. Geriatr. Rep.* 8, 55–66. doi: 10.1007/s13670-019-0275-3
15. Stephens, J., and Allen, J. (2013). Mobile phone interventions to increase physical activity and reduce weight: a systematic review. *J. Cardiovasc. Nurs.* 28, 320–329. doi: 10.1097/JCN.0b013e318250a3e7
16. Lau, P. W. C., Lau, E. Y., Wong, D. P., and Ransdell, L. (2011). A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. *J. Med. Internet Res.* 13:e48. doi: 10.2196/jmir.1533
17. Tong, H. L., and Laranjo, L. (2018). The use of social features in mobile health interventions to promote physical activity: a systematic review. *NPJ Digit. Med.* 1:43. doi: 10.1038/s41746-018-0051-3
18. Hardeman, W., Houghton, J., Lane, K., Jones, A., and Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *Int. J. Behav. Nutr. Phys. Act.* 16:31. doi: 10.1186/s12966-019-0792-7
19. Ehn, M., Eriksson, L. C., Åkerberg, N., and Johansson, A. C. (2018). Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. *JMIR Mhealth Uhealth* 6:e34. doi: 10.2196/mhealth.8345
20. Hamasaki, H. (2018). *Efficacy of Wearable Devices to Measure and Promote Physical Activity in the Management of Diabetes*. Retrieved from: https://emj.europeanmedical-group.com/wp-content/uploads/sites/2/2018/11/Efficacy-of-Wearable-Devices-to-Measure-and-Promote-Physical-Activity-in-the-Management-of-Diabetes.pdf
21. Bort-Roig, J., Gilson, N. D., Puig-Ribera, A., Contreras, R. S., and Trost, S. G. (2014). Measuring and influencing physical activity with smartphone technology: a systematic review. *Sports Med.* 44, 671–686. doi: 10.1007/s40279-014-0142-5
22. Matthews, J., Win, K. T., Oinas-Kukkonen, H., and Freeman, M. (2016). Persuasive technology in mobile applications promoting physical activity: a systematic review. *J. Med. Syst.* 40:72. doi: 10.1007/s10916-015-0425-x
23. Ghanvatkar, S., Kankanhalli, A., and Rajan, V. (2019). User models for personalized physical activity interventions: scoping review. *J. Med. Internet Res.* 7:e11098. doi: 10.2196/preprints.11098
24. Almutari, N., and Orji, R. (2019). "How effective are social influence strategies in persuasive apps for promoting physical activity? A systematic review," in *Adjunct Publication of the 27th Conference* (New York, NY), 167–172. doi: 10.1145/3314183.3323855
25. Win, K. T. M., Roberts, R. H., and Oinas-Kukkonen, H. (2018). Persuasive system features in computer-mediated lifestyle modification interventions for physical activity. *Inform. Health Soc. Care* 44, 376–404. doi: 10.1080/17538157.2018.1511565
26. Win, K. T., Ramaprasad, A., and Syn, T. (2019). Ontological review of persuasion support systems (PSS) for health behavior change through physical activity. *J. Med. Syst.* 43:49. doi: 10.1007/s10916-019-1159-y
27. Prince, S. A., Saunders, T. J., Gresty, K., and Reid, R. D. (2014). A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. *Obes. Rev.* 15, 905–919. doi: 10.1111/obr.12215
28. Schembre, S. M., Liao, Y., Robertson, M. C., Dunton, G. F., Kerr, J., Haffey, M. E., et al. (2018). Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. *J. Med. Internet Res.* 20:e106. doi: 10.2196/jmir.8701
29. Schoeppe, S., Alley, S., Lippevelde, W. V., Bray, N. A., Williams, S. L., Duncan, M. J., et al. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *Int. J. Behav. Nutr. Phys. Act.* 13:127. doi: 10.1186/s12966-016-0454-y
30. Yim, J., and Graham, T. C. N. (2007). "Using games to increase exercise motivation," in *Proceedings of the 2007 Conference on Future Play–Future Play '07* (New York, NY). doi: 10.1145/1328202.1328232
31. Lister, C., West, J. H., Cannon, B., Sax, T., and Brodegard, D. (2014). Just a fad? Gamification in health and fitness apps. *JMIR Serious Games.* 2:e9. doi: 10.2196/games.3413
32. Fogg, B. J. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco, CA: Ubiquity, Morgan Kaufmann.
33. Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., J., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J. Clin. Epidemiol.* 62, e1–e34. doi: 10.1016/j.jclinepi.2009.06.006
34. Mueller, F., Agamanolis, S., and Picard, R. (2003). Exertion interfaces: sports over a distance for social bonding and fun. *Proc. Conf. Hum. Fact. Comput. Syst.* 5, 561–568. doi: 10.1145/642611.642709
35. Jafarinaimi, N., Forlizzi, J., Hurst, A., and Zimmerman, J. (2004). "Breakaway: an ambient display designed to change human behavior," in *BT–In Proceedings of the Conference on Computer-Human Interaction* (New York, NY), 1945–1948. doi: 10.1145/1056808.1057063
36. Nakamura, S., Minakuchi, M., and Tanaka, K. (2005). "Energy browser: to make exercise enjoyable and interesting," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (New York, NY). doi: 10.1145/1178477.1178521
37. Eyck, A., Geerlings, K., Karimova, D., Meerbeek, B., Wang, L., Usselsteijn, W., et al. (2006). "Effect of a virtual coach on athletes' motivation," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (Berlin; Heidelberg: Springer Verlag) 158–161. doi: 10.1007/11755494_22
38. Gasser, R., Brodbeck, D., Degen, M., Luthiger, J., Wyss, R., and Reichlin, S. (2006). "Persuasiveness of a mobile lifestyle coaching application using social facilitation," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3962 LNCS*, eds W. A. IJsselsteijn, Y. A. W. de Kort, C. Midden, B. Eggen, and E. van den Hoven (Berlin; Heidelberg: Springer), 27–38. doi: 10.1007/11755494_5
39. Lin, J. J., Mamykina, L., Strub, H. B., Lindtner, S., and Delajoux, G. (2006). "Fish'n'Steps: encouraging physical activity with an interactive computer game," in *UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006* (Orange County, CA), 261–278. doi: 10.1007/11853565_16
40. McCreadie, C., Raper, J., Gunesh, A., Wood, J., Carey, K., Petrie, H., et al. (2006). "Persuasive technology for leisure and health: Development of a personal navigation tool," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3962 LNCS*, eds W. A. IJsselsteijn, Y. A. W. de Kort, C. Midden, B. Eggen, and E. van den Hoven (Berlin; Heidelberg: Springer), 187–190. doi: 10.1007/11755494_27
41. Toscos, T., Faber, A., An, S., and Gandhi, M. P. (2006). "Chick clique–persuasive technology to motivate teenage girls to exercise," in *Extended Abstracts Proceedings of the 2006 Conference on Human Factors in Computing Systems, CHI 2006* (Montréal, QC). doi: 10.1145/1125451.1125805
42. Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., et al. (2007). Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. *Mob. Netw. App.* 12, 185–199. doi: 10.1007/s11036-007-0011-7
43. Sohn, M., and Lee, J. (2007). "UP health: ubiquitously persuasive health promotion with an instant messaging system," in *25th SIGCHI Conference on Human Factors in Computing Systems 2007, CHI 2007* (New York, NY), 2663–2668.
44. Consolvo, S., and Klasnja, P. (2008). "Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays," in *UbiComp '08: Proceedings of the 10th International Conference on Ubiquitous Computing* (New York, NY), 54–63. doi: 10.1145/1409635.1409644
45. De Oliveira, R., and Oliver, N. (2008). "Triplebeat: enhancing exercise performance with persuasion," in *Proceedings of the 10th Conference on Human-Computer Interaction with Mobile Devices and Services, Mobile HCI 2008* (Amsterdam) 255–264. doi: 10.1145/1409240.1409268
46. Fujiki, Y., Kazakos, K., Puri, C., Buddharaju, P., Pavlidis, I., and Levine, J. (2008). NEAT-o-games: blending physical activity and fun in the daily routine. *ACM Comput. Entertain.* 6:21. doi: 10.1145/1371216.1371224
47. Fujinami, K., and Riekki, J. (2008). "A case study on an ambient display as a persuasive medium for exercise awareness," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, eds H. Oinas-Kukkonen, P. Hasle, M.

Harjumaa, K. Segerståhl, and P. Øhrstrøm (Berlin; Heidelberg: Springer Verlag), 266–269. doi: 10.1007/978-3-540-68504-3_28

48. Gasca, E., Favela, J., and Tentori, M. (2008). "Persuasive virtual communities to promote a healthy lifestyle among patients with chronic diseases," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5411 LNCS*, eds R. O. Briggs, P. Antunes, G. J. de Vreede, and A. S. Read (Berlin; Heidelberg: Springer), 74–82. doi: 10.1007/978-3-540-92831-7_7
49. Obermair, C., Reitberger, W., Meschtscherjakov, A., Lankes, M., and Tscheligi, M. (2008). "PerFrames: persuasive picture frames for proper posture," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, eds H. Oinas-Kukkonen, P. Hasle, M. Harjumaa, K. Segerståhl, and P. Øhrstrøm (Berlin; Heidelberg: Springer Verlag), 128–139. doi: 10.1007/978-3-540-68504-3_12
50. Spruijt-Metz, D., Nguyen-Michel, S. T., Goran, M. I., Chou, C. P., and Huang, T. T. K. (2008). Reducing sedentary behavior in minority girls via a theory-based, tailored classroom media intervention. *Int. J. Pediatr. Obes.* 3, 240–248. doi: 10.1080/17477160802113415
51. Toscos, T., Faber, A., Connelly, K., and Upoma, A. M. (2008). "Encouraging physical activity in teens can technology help reduce barriers to physical activity in adolescent girls?" in *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare 2008, PervasiveHealth* (Tampere). doi: 10.1109/PCTHEALTH.2008.4571073
52. Albaina, I. M., Visser, T., van der Mast, C. A. P. G., and Vastenburg, M. H. (2009). "Flowie: a persuasive virtual coach to motivate elderly individuals to walk," in *3rd International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth 2009* (London). doi: 10.4108/ICST.PERVASIVEHEALTH2009.5949
53. Fialho, A., and Van Den Heuvel, H. (2009). "ActiveShare: sharing challenges to increase physical activity," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Extended Abstracts Volume* (Boston, MA), 4159–4164. doi: 10.1145/1520340.1520633
54. Gerber, B. S., Stolley, M. R., Thompson, A. L., Sharp, L. K., and Fitzgibbon, M. L. (2009). Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. *Health Informatics J.* 15, 17–25. doi: 10.1177/1460458208099865
55. Harjumaa, M., and Oinas-kukkonen, H. (2009). "Understanding persuasive software functionality in practice: a field trial of polar FT60," in *Persuasive Technology, Fourth International Conference, PERSUASIVE 2009* (Claremont, CA). doi: 10.1145/1541948.1541952
56. Jung, Y., Li, K. J., Janissa, N. S. W., Gladys, L. C., and Lee, K. M. (2009). "Games for a better life: Effects of playing wii games on the well-being of seniors in a long-term care facility," in *BT—6th Australasian Conference on Interactive Entertainment, IE 2009* (New York, NY: University of Technology Sydney; ACM Special Inter.).
57. Lacroix, J., Saini, P., and Goris, A. (2009). "Understanding user cognitions to guide the tailoring of persuasive technology-based physical activity interventions," in *Persuasive Technology, Fourth International Conference, PERSUASIVE 2009* (Claremont, CA), 1. doi: 10.1145/1541948.1541961
58. Mueller, F., Gibbs, M. R., and Vetere, F. (2009). "Design influence on social play in distributed exertion games," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009* (Boston, MA). doi: 10.1145/1518701.1518938
59. Otten, J. J., Jones, K. E., Littenberg, B., and Harvey-Berino, J. (2009). Effects of television viewing reduction on energy intake and expenditure in overweight and obese adults: a randomized controlled trial. *Archiv. Intern. Med.* 169, 2109–2115. doi: 10.1001/archinternmed.2009.430
60. Arteaga, S. M., Kudeki, M., Woodworth, A., and Kurniawan, S. (2010). "Mobile system to motivate teenagers' physical activity," in *IDC '10: Proceedings of the 9th International Conference on Interaction Design and Children* (New York, NY), 1–10. doi: 10.1145/1810543.1810545
61. Berkovsky, S., Coombe, M., and Helmer, R. (2010). "Activity interface for physical activity motivating games," in *IUI '10: Proceedings of the 15th International Conference on Intelligent User Interfaces* (New York, NY), 273–276. doi: 10.1145/1719970.1720009
62. Bieber, G., Koldrack, P., Sablowski, C., Peter, C., and Urban, B. (2010). "Mobile physical activity recognition of stand-up and sit-down transitions for user behavior analysis," in *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments–PETRA '10* (New York, NY). doi: 10.1145/1839294.1839354
63. Foster, D., Linehan, C., Kirman, B., Lawson, S., and James, G. (2010). "Motivating physical activity at work: using persuasive social media for competitive step counting," in *Proceedings of the 14th International Academic MindTrek Conference on Envisioning Future Media Environments–MindTrek '10* (New York, NY), 111. doi: 10.1145/1930488.1930510
64. Kientz, J. A., Choe, E. K., Birch, B., Maharaj, R., Fonville, A., Glasson, C., et al. (2010). "Heuristic evaluation of persuasive health technologies," in *ACM International Health Informatics Symposium, IHI 2010* (Arlington, VA). doi: 10.1145/1882992.1883084
65. Munguía, A., and Santana, P. C. (2010). "A wearable augmented reality system to improve the quality of life of sedentary people," in *Proceedings of the 3rd Mexican Workshop on Human Computer Interaction*, 8–11. Retrieved from: http://www.pedrosantana.mx/papers/MexIHC-2010_augmented-reality.pdf
66. Penados, A. L., Gielen, M., Stappers, P. J., and Jongert, T. (2010). Get up and move: an interactive cuddly toy that stimulates physical activity. *Pers. Ubiquit. Comput.* 14, 397–406. doi: 10.1007/s00779-009-0270-3
67. Stanley, K. G., Livingston, I., Bandurka, A., Kapiszka, R., and Mandryk, R. L. (2010). "PiNiZoRo: a GPS-based exercise game for families," in *International Academic Conference on the Future of Game Design and Technology, Future Play* (New York, NY), 243–246. doi: 10.1145/1920778.1920817
68. Young, M. M. (2010). "Twitter me: using micro-blogging to motivate teenagers to exercise," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6105 LNCS*, eds R. Winter, J. L. Zhao, and S. Aier (Berlin; Heidelberg: Springer), 439–448. doi: 10.1007/978-3-642-13335-0_30
69. Lim, B. Y., Shick, A., Harrison, C., and Hudson, S. E. (2011). "Pediluma: motivating physical activity through contextual information and social influence," in *Proceedings of the 5th International Conference on Tangible and Embedded Interaction 2011* (Funchal), 173–180. Retrieved from: http://dl.acm.org/citation.cfm?id=1935736
70. Lin, Y., Jessurun, J., de Vries, B., and Timmermans, H. (2011). "Motivate: towards context-aware recommendation mobile system for healthy living," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, 250–253. doi: 10.4108/icst.pervasivehealth.2011.246030
71. Sakai, R., Peteghem, S. V. L., Van De Sande, Banach, P., and Kaptein, M. (2011). "Personalized persuasion in ambient intelligence: the APStairs system," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7040 LNCS*, eds D. V. Keyson, M. L. Maher, N. Streitz, A. Cheok, J. C. Augusto, R. Wichert, G. Englebienne, H. Aghajan, and B. J. A. Kröse (Berlin; Heidelberg: Springer), 205–209. doi: 10.1007/978-3-642-25167-2_26
72. Al-Hrathi, R., Karime, A., Al-Osman, H., and El Saddik, A. (2012). "Exerlearn bike: an exergaming system for children's educational and physical well-being," in *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2012* (Melbourne, VIC), 489–494. doi: 10.1109/ICMEW.2012.91
73. Alkhajah, T. A., Reeves, M. M., Eakin, E. G., Winkler, E. A. H., Owen, N., Healy, G. N., et al. (2012). Sit-stand workstations: a pilot intervention to reduce office sitting time. *Am. J. Prev. Med.* 43, 298–303. doi: 10.1016/j.amepre.2012.05.027
74. Ayubi, S. U., and Parmanto, B. (2012). "PersonA: persuasive social network for physical activity," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (San Diego, CA), 2153–2157. doi: 10.1109/EMBC.2012.6346387
75. Berkovsky, S., Freyne, J., and Coombe, M. (2012). "Physical activity motivating games: be active and get your own reward," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (ACM Transactions on Computer-Human Interaction), 1–41. doi: 10.1145/2395131.2395139
76. Chittaro, L., and Sioni, R. (2012). "Turning the classic snake mobile game into a location-based exergame that encourages walking," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7284 LNCS*,

eds M. Bang and E. L. Ragnemalm (Berlin; Heidelberg: Springer), 43–54. doi: 10.1007/978-3-642-31037-9_4

77. Dewa, C. S., deRuiter, W., Chau, N., and Karioja, K. (2012). Walking for wellness: using pedometers to decrease sedentary behaviour and promote mental health. *Int. J. Mental Health Promot.* 11, 24–28. doi: 10.1080/14623730.2009.9721784
78. Dobbins, C., Fergus, P., Merabti, M., and Llewellyn-Jones, D. (2012). "Monitoring and measuring sedentary behaviour with the aid of human digital memories," in *2012 IEEE Consumer Communications and Networking Conference, CCNC'2012* (Las Vegas, NV). doi: 10.1109/ICC.2012.6364905
79. Evans, R. E., Fawole, H. O., Sheriff, S. A., Dall, P. M., Grant, P. M., and Ryan, C. G. (2012). Point-of-choice prompts to reduce sitting time at work: a randomized trial. *Am. J. Prev. Med.* 43, 293–297. doi: 10.1016/j.amepre.2012.05.010
80. Fan, C., Forlizzi, J., and Dey, A. K. (2012). "A spark of activity: exploring information art as visualization for physical activity," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing–UbiComp '12* (New York, NY), 81. doi: 10.1145/2370216.2370229
81. Grosinger, J., Vetere, F., and Fitzpatrick, G. (2012). "Agile life : addressing knowledge and social motivations for active aging," in *Proceedings of the 24th Australian Computer-Human Interaction Conference* (New York, NY). doi: 10.1145/2414536.2414566
82. Hamiliton, I., Imperatore, G., and Dunlop, M. D. (2012). "Walk2Build: a GPS game for mobile exergaming with city visualization," in *MobileHCI '12: Proceedings of the 14th International Conference on Human-Computer Interaction With Mobile Devices and Services Companion* (New York, NY), 17–22. Retrieved from: http://dl.acm.org/citation.cfm?id=2371670
83. Karime, A., Hafidh, B., Khaldi, A., Aljaam, J. M., and El Saddik, A. (2012). MeMaPads: "Enhancing children's well-being through a physically interactive memory and math games," in *2012 IEEE I2MTC–International Instrumentation and Measurement Technology Conference, Proceedings* (Graz), 2563–2566. doi: 10.1109/I2MTC.2012.6229520
84. Munson, S., and Consolvo, S. (2012). "Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity," in *6th International Conference on Pervasive Computing Technologies for Healthcare* (San Diego, CA). doi: 10.4108/icst.pervasivehealth.2012.248691
85. Mutsuddi, A., and Connelly, K. (2012). "Text messages for encouraging physical activity are they effective after the novelty effect wears off?" in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops* (San Diego, CA), 33–40. doi: 10.4108/icst.pervasivehealth.2012.248715
86. Park, T., Song, J., Hwang, I., Lee, U., Lee, S. I., Yoo, C., et al. (2012). "ExerLink: enabling pervasive social exergames with heterogeneous exercise devices," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services–MobiSys '12* (New York, NY), 15. doi: 10.1145/2307636.2307639
87. Pina, L., Ramirez, E., and Griswold, W. (2012). "Fitbit+: a behavior-based intervention system to reduce sedentary behavior," in *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)* (San Diego, CA). doi: 10.4108/icst.pervasivehealth.2012.248761
88. Probst, K., Perteneder, F., Jakob, L., Haller, M., Schrempf, A., and Glöckl, J. (2012). "Active office: towards an activity–promoting office workplace design," in *Extended Abstracts on Human Factors in Computing Systems (CHI'12)* (New York, NY), 2165–2170. doi: 10.1145/2212776.2223770
89. Pronk, N. P., Katz, A. S., Lowry, M., and Payfer, J. R. (2012). Reducing occupational sitting time and improving worker health: the take-a-stand project (2011). *Prev. Chron. Dis.* 9, 1–9. doi: 10.5888/pcd9.110323
90. Ramanathan, N., Alquaddoomi, F., Falaki, H., George, D. C., Hsieh, K., Jenkins, J., et al. (2012). "ohmage: an open mobile system for activity and experience sampling," in *6th International Conference on Pervasive Computing Technologies for Healthcare* (San Diego, CA). doi: 10.4108/icst.pervasivehealth.2012.248705
91. Zwinderman, M. J., Shirzad, A., Ma, X., Bajracharya, P., Sandberg, H., and Kaptein, M. C. (2012). "Phone row: a smartphone game designed to persuade people to engage in moderate-intensity physical activity," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7284 LNCS*, eds M. Bang and E. L. Ragnemalm (Berlin; Heidelberg: Springer), 55–66. doi: 10.1007/978-3-642-31037-9_5
92. Barwais, F. A., Cuddihy, T. F., and Tomson, L. M. (2013). Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial. *Health Qual. Life Outcomes* 11:183. doi: 10.1186/1477-7525-11-183
93. Carr, L. J., Karvinen, K., Peavler, M., Smith, R., and Cangelosi, K. (2013). Multicomponent intervention to reduce daily sedentary time: a randomised controlled trial. *BMJ Open* 3:10. doi: 10.1136/bmjopen-2013-003261
94. Cercos, R., and Mueller, F. F. (2013). Watch your steps: designing a semi-public display to promote physical activity. *IE* 2013, 2:1–2:6. doi: 10.1145/2513002.2513016
95. Fortmann, J., Stratmann, T., Boll, S., Poppinga, B., and Heuten, W. (2013). "Make me move at work! An ambient light display to increase physical activity," in *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops* (Venice). doi: 10.4108/icst.pervasivehealth.2013.252089
96. Gilson, N. D., Faulkner, G., Murphy, M. H. M., Meyer, R. U., Washington, T., et al. (2013). Walk@Work: an automated intervention to increase walking in university employees not achieving 10,000 daily steps. *Prev. Med.* 56, 283–287. doi: 10.1016/j.ypmed.2013.01.022
97. Glynn, L. G., Hayes, P. S., Casey, M., Glynn, F., Alvarez-Iglesias, A., Newell, J., et al. (2013). SMART MOVE–a smartphone-based intervention to promote physical activity in primary care: Study protocol for a randomized controlled trial. *Trials* 14:157. doi: 10.1186/1745-6215-14-157
98. Healy, G. N., Eakin, E. G., LaMontagne, A. D., Owen, N., E., Winkler, A. H., et al. (2013). Reducing sitting time in office workers: short-term efficacy of a multicomponent intervention. *Prev. Med.* 57, 43–48. doi: 10.1016/j.ypmed.2013.04.004
99. Hirano, S. H., Farrell, R. G., Danis, C. M., and Kellogg, W. A. (2013). "WalkMinder: encouraging an active lifestyle using mobile phone interruptions," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (New York, NY), 1431–1436. doi: 10.1145/2468356.2468611
100. Hong, Y., Dahlke, D. V., Ory, M., Hochhalter, A., Reynolds, J., Purcell, N. P., et al. (2013). Designing icanfit: a mobile-enabled web application to promote physical activity for older cancer survivors. *JMIR Res. Protoc.* 2:e12. doi: 10.2196/resprot.2440
101. Jago, R., Sebire, S. J., Turner, K. M., Bentley, G. F., Goodred, J. K., Fox, K. R., et al. (2013). Feasibility trial evaluation of a physical activity and screen-viewing course for parents of 6 to 8 year-old children: teamplay. *Int. J. Behav. Nutr. Phys. Act.* 10:31. doi: 10.1186/1479-5868-10-31
102. Keung, C., Lee, A., O'Keefe, M., and Lu, S. (2013). "BunnyBolt: a mobile fitness app for youth," in *IDC '13: Proceedings of the 12th International Conference on Interaction Design and Children* (New York, NY), 585–588. doi: 10.1145/2485760.2485871
103. Khalil, A., and Abdallah, S. (2013). Harnessing social dynamics through persuasive technology to promote healthier lifestyle. *Comput. Hum. Behav.* 29, 2674–2681. doi: 10.1016/j.chb.2013.07.008
104. Macvean, A., and Robertson, J. (2013). "Understanding exergame users' physical activity, motivation and behavior over time," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY). doi: 10.1145/2470654.2466163
105. Marinac, C., Merchant, G., Godbole, S., Chen, J., Kerr, J., Clark, B., et al. (2013). "The feasibility of using SenseCams to measure the type and context of daily sedentary behaviors," in *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference* (New York, NY). doi: 10.1145/2526667.2526674
106. McMahon, S., Vankipuram, M., and Fleury, J. (2013). Mobile computer application for promoting physical activity. *J. Gerontol. Nurs.* 39, 15–20. doi: 10.3928/00989134-20130226-01
107. Nakajima, T., and Lehdonvirta, V. (2013). Designing motivation using persuasive ambient mirrors. *Pers. Ubiquit. Comput.* 17, 107–126. doi: 10.1007/s00779-011-0469-y
108. Park, T., Lee, U., Lee, B., Lee, H., Son, S., Song, S., et al. (2013). "ExerSync: facilitating interpersonal synchrony in social exergames," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work–CSCW '13* (New York, NY), 409. doi: 10.1145/2441776.2441823

109. Parry, S., and Straker, L. (2013). The contribution of office work to sedentary behaviour associated risk. *BMC Public Health* 13:1. doi: 10.1186/1471-2458-13-296
110. Peeters, M., Brombacher, A., Megens, C., van den Hoven, E., and Hummels, C. (2013). "Social Stairs: taking the Piano Staircase towards long- term behavioral change," in *Proceedings of the 8th international conference on Persuasive Technology*, 1–6. Retrieved from: https://link.springer.com/content/pdf/10.1007/978-3-642-37157-8_21.pdf
111. Probst, K., Lindlbauer, D., and Greindl, P. (2013). "Rotating, tilting, bouncing: using an interactive chair to promote activity in office environments," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (New York, NY), 79–84. doi: 10.1145/2468356.2468372
112. Rodríguez, M. D., Roa, J. R., Morán, A. L., and Nava-Muñoz, S. (2013). CAMMInA: a mobile ambient information system to motivate elders to exercise. *Pers. Ubiquit. Comput.* 17, 1127–1134. doi: 10.1007/s00779-012-0561-y
113. Van Dantzig, S., Geleijnse, G., and Van Halteren, A. T. (2013). Toward a persuasive mobile application to reduce sedentary behavior. *Pers. Ubiquit. Comput* (New York, NY), 17, 1237–1246. doi: 10.1007/s00779-012-0588-0
114. Bond, D. S., Thomas, J. G., Raynor, H. A., Moon, J., Sieling, J., Trautvetter, J., et al. (2014). B-MOBILE–a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. *PLoS ONE* 9:e100821. doi: 10.1371/journal.pone.0100821
115. Chen, Y., and Pu, P. (2014). "HealthyTogether: exploring social incentives for mobile fitness applications," in *2nd International Symposium of Chinese CHI, Chinese CHI 2014* (New York, NY), 25–34. doi: 10.1145/2592235.2592240
116. Chen, Y. X., Chiang, S. S., Chih, S. Y., Liao, W. C., Lin, S. Y., Yang, S. H., et al. (2014). "Opportunities for persuasive technology to motivate heavy computer users for stretching exercise," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8462 LNCS*, eds A. Spagnolli, L. Chittaro, and L. Gamberini (Cham: Springer Verlag), 25–30. doi: 10.1007/978-3-319-07127-5_3
117. Clinkenbeard, D., Clinkenbeard, J., Faddoul, G., Kang, H., Mayes, S., Toygar, A., et al. (2014). "What's your 2%? A pilot study for encouraging physical activity using persuasive video and social media," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (New York, NY). doi: 10.1007/978-3-319-07127-5_5
118. Dobbins, C., Merabti, M., Fergus, P., and Llewellyn-Jones, D. (2014). "A user-centred approach to reducing sedentary behaviour," in *11th Annual IEEE Consumer Communications & Networking Conference (CCNC'14)* (Las Vegas, NV), 1–6. doi: 10.1109/CCNC.2014.6866645
119. Far, I. K., Ibarra, F., Baez, M., and Casati, F. (2014). *Virtual Social Gym: a Persuasive Training Platform for Independently Living Seniors*. Retrieved from: http://demo.socialgym.org/
120. Ferreira, M. J., Caraban, A. K., and Karapanos, E. (2014). "Breakout: predicting and breaking sedentary behaviour at work," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 2407–2412. Retrieved from: https://dl.acm.org/citation.cfm?id=2581330
121. Fritz, T., Huang, E. M., and Murphy, G. C. (2014). "Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness," in *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY), 241–244. doi: 10.1145/2556288.2557383
122. Gao, Y., Gerling, K. M., Mandryk, R. L., and Stanley, K. G. (2014). "Decreasing sedentary behaviours in pre-adolescents using casual exergames at school," in *CHI PLAY '14: Proceedings of the first ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play* (New York, NY), 97–106. doi: 10.1145/2658537.2658693
123. He, Q., and Agu, E. (2014). "On11: an activity recommendation application to mitigate sedentary lifestyle," in *Proceedings of the 2014 Workshop on Physical Analytics* (New York, NY), 3–8.
124. Lane, N. D., Lin, M., Mohammod, M., Yang, X., Lu, H., Cardone, G., et al. (2014). BeWell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mob. Netw. Appl.* 19, 345–359. doi: 10.1007/s11036-013-0484-5
125. Moghimi, M., Wu, W., Chen, J., and Godbole, S. (2014). "Analyzing sedentary behavior in life-logging images," in *International Conference on Image Processing (ICIP)* (Paris), 1011–1015. doi: 10.1109/ICIP.2014.7025202
126. Neuhaus, M., Healy, G. N., Dunstan, D. W., Owen, N., and Eakin, E. G. (2014). Workplace sitting and height-adjustable workstations: a randomized controlled trial. *Am. J. Prev. Med.* 46, 30–40. doi: 10.1016/j.amepre.2013.09.009
127. Sun, M., Burke, L. E., Mao, Z., Chen, Y. H., Chen, C., Bai, Y., et al. (2014). "eButton : a wearable computer for health monitoring and personal assistance," in *Design Automation Conference (DAC), 2014 51st ACM/EDAC/IEEE* (New York, NY), 1–6. doi: 10.1145/2593069.2596678
128. Zaman, K. I., Yli-Piipari, S., and Hnat, T. W. (2014). "Kinematic-based sedentary and light-intensity activity detection for wearable medical applications," in *MMA '14: Proceedings of the 1st Workshop on Mobile Medical Applications* (New York, NY), 28–33. doi: 10.1145/2676431.2676433
129. Zuckerman, O., and Gal-Oz, A. (2014). Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. *Pers. Ubiquit. Comput.* 18, 1705–1719. doi: 10.1007/s00779-014-0783-2
130. Donath, L., Faude, O., Schefer, Y., Roth, R., and Zahner, L. (2015). Repetitive daily point of choice prompts and occupational sit-stand transfers, concentration and neuromuscular performance in office workers: an RCT. *Int. J. Environ. Res. Public Health* 12, 4340–4353. doi: 10.3390/ijerph120404340
131. Hao, T., Xing, G., and Zhou, G. (2015). "RunBuddy: a smartphone system for running rhythm monitoring," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing–UbiComp '15* (New York, NY), 133–144. doi: 10.1145/2750858.2804293
132. Bexheti, A., Fedosov, A., Findahl, J., Langheinrich, M., and Niforatos, E. (2015). "Re-live the moment: visualizing run experiences to motivate future exercises," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct–MobileHCI '15* (New York, NY), 986–993. doi: 10.1145/2786567.2794316
133. Braun, A., Schembri, I., and Frank, S. (2015). "Exerseat–sensor-supported exercise system for ergonomic microbreaks," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, eds B. De Ruyter, A. Kameas, P. Chatzimisios, and I. Mavrommati (Cham: Springer Verlag), 236–251. doi: 10.1007/978-3-319-26005-1_16
134. Chatta, A., Hurst, T., Samaraweera, G., Guo, R., and Quarles, J. (2015). "Get off the couch get off the couch: an approach to utilize sedentary commercial games as exergames," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play–CHI PLAY '15* (New York, NY), 47–56. doi: 10.1145/2793107.2793115
135. Gouveia, R., Karapanos, E., and Hassenzahl, M. (2015). "How do we engage with activity trackers? A longitudinal study of Habito," in *UbiComp '15* (New York, NY). doi: 10.1145/2750858.2804290
136. Gupta, S., and Sood, S. (2015). "Context aware mobile agent for reducing stress and obesity by motivating physical activity: a design approach," in *International Conference on Computing for Sustainable Global Development, INDIACom 2015*, 962–966. Retrieved from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84960868913&partnerID=40&md5=73862401886e0bda300f58d8b8603d97
137. Mansart, C., Sukitphittayanon, S., Pantongkhum, P., and Thaicharoen, S. (2015). "Go run go: an android game-story application for aiding motivation to exercise," in *Proceedings—2015 IEEE International Symposium on Multimedia, ISM 2015* (Miami, FL), 407–410. doi: 10.1109/ISM.2015.49
138. Min, D. A., Kim, K. Y., Kim, Y. S., Jung, E., Jang, S. A., and Lee, J. H. (2015). "Pretty pelvis: a virtual pet application that breaks sedentary time by promoting gestural interaction," in *Crossing: Proceedings of CHI '15 Extended Abstracts on Human Factors in Computing Systems* (New York, NY), 1259–1264. doi: 10.1145/2702613.2732807
139. Ornelas, T., Caraban, A., Gouveia, R., and Karapanos, E. (2015). "CrowdWalk: leveraging the wisdom of the crowd to inspire walking activities," in *Adjunct Proceedings of Ubicomp'15* (New York, NY), 213–216. doi: 10.1145/2800835.2800923
140. Pellegrini, C. A., Hoffman, S. A., Daly, E. R., Murillo, M., Iakovlev, G., and Spring, B. (2015). Acceptability of smartphone technology to interrupt

sedentary time in adults with diabetes. *Transl. Behav. Med.* 5, 307–314. doi: 10.1007/s13142-015-0314-3
141. Spiesberger, P., Jungwirth, F., Wöss, C., Bachl, S., Harms, J., and Grechenig, T. (2015). "Woody: a location-based smartphone game to increase children's outdoor activities in urban environments," in *ACM International Conference Proceeding Series* (New York, NY). doi: 10.1145/2836041.2841210
142. Spinney, R., Smith, L., Ucci, M., Fisher, A., Konstantatou, M., Sawyer, A., et al. (2015). Indoor tracking to understand physical activity and sedentary behaviour: exploratory study in UK office buildings. *PLoS ONE* 10:e127688. doi: 10.1371/journal.pone.0127688
143. Stragier, J., Evens, T., and Mechant, P. (2015). *Broadcast Yourself: An Exploratory Study of Sharing Physical Activity on Social Networking Sites*. Media International Australia.
144. Van Almkerk, M., Bierling, B. L., Leermakers, N., Vinken, J., and Timmermans, A. A. A. (2015). "Improving posture and sitting behavior through tactile and visual feedback in a sedentary environment," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (Milan), 4570–4573. doi: 10.1109/EMBC.2015.7319411
145. van Schagen, J., Gribnau, M., de Leeuw, J., Los, B., Cleintuar, N., and Bidarra, R. (2015). "Super starfish mania: fish for friends," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play–CHI PLAY '15* (New York, NY), 787–790. doi: 10.1145/2793107.2810275
146. Puig-Ribera, A., Martínez-Lemos, I., Milà, R., Giné-Garriga, M., Gilson, N. D., Bort-Roig, J., et al. (2016). Patterns of impact resulting from a 'sit less, move more' web-based program in sedentary office employees. *PLoS ONE* 10:e0122474. doi: 10.1371/journal.pone.0122474
147. Brakenridge, C. L., Fjeldsoe, B. S., Young, D. C., Winkler, E. A. H., Dunstan, D. W., Healy, G. N., et al. (2016). Evaluating the effectiveness of organisational-level strategies with or without an activity tracker to reduce office workers' sitting time: a cluster-randomised trial. *Int. J. Behav. Nutr. Phys. Act.* 13, 1–15. doi: 10.1186/s12966-016-0441-3
148. Buddharaju, P., and Lokanathan, Y. (2016). "Mobile exergaming: exergames on the go," in *Proceedings of the International Workshop on Mobile Software Engineering and Systems–MOBILESoft* (New York, NY), 16, 25–26. doi: 10.1145/2897073.2897125
149. De Cocker, K., De Bourdeaudhuij, I., Cardon, G., and Vandelanotte, C. (2016). The effectiveness of a web-based computer-tailored intervention on workplace sitting: a randomized controlled trial. *J. Med. Internet Res.* 18, 1–14. doi: 10.2196/jmir.5266
150. Dharia, S., Jain, V., Patel, J., Vora, J., Yamauchi, R., Eirinaki, M., et al. (2016). "PRO-fit: exercise with friends," in *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016* (San Francisco, CA), 1430–1433. doi: 10.1109/ASONAM.2016.7752437
151. Gilson, N. D., Ng, N., Pavey, T. G., Ryde, G. C., Straker, L., and Brown, W. J. (2016). Project Energise: using participatory approaches and real time computer prompts to reduce occupational sitting and increase work time physical activity in office workers. *J. Sci. Med. Sport* 19, 926–930. doi: 10.1016/j.jsams.2016.01.009
152. Hadgraft, N. T., Healy, G. N., Owen, N. E., Winkler, A. H., Lynch, B. M., et al. (2016). Office workers' objectively assessed total and prolonged sitting time: individual-level correlates and worksite variations. *Prev. Med. Rep.* 4, 184–191. doi: 10.1016/j.pmedr.2016.06.011
153. Haque, Md. S., Abdullah, W. M., Rahaman, S., Kangas, M., and Jämsä, T. (2016). "Persuasive health and wellbeing application: a theory-driven design in promoting physical activity," in *1st International Conference on Medical Engineering, Health Informatics and Technology, MediTec 2016* (Dhaka). doi: 10.1109/MEDITEC.2016.7835369
154. Huang, Y. (2016). "How to design internet of things to encourage office workers to take more regular micro-breaks," in *ECCE '16: The European Conference on Cognitive Ergonomics* (Nottingham), 1–3. doi: 10.1145/2970930.2970963
155. Kim, J., Lee, N. H. B., Bae, C., and Cho, J. D. (2016). "A feedback system for the prevention of forward head posture in sedentary work environments," in *The 2016 ACM Conference Companion Publication* (New York, NY), 161–164. doi: 10.1145/2908805.2909414
156. Kumar, R., Bayliff, A., De, D., Evans, A., Das, S. K., and Makos, M. (2016). "Care-chair: sedentary activities and behavior assessment with smart sensing on chair backrest," in *2016 IEEE International Conference on Smart Computing, SMARTCOMP* (St. Louis, MO) 2016, 1–8. doi: 10.1109/SMARTCOMP.2016.7501682
157. Mohadis, H. M., and Ali, N. M. (2016). "Designing persuasive application to encourage physical activity at workplace among older workers," in *2016 6th International Conference on Digital Information and Communication Technology and Its Applications, DICTAP 2016* (Konya). doi: 10.1109/DICTAP.2016.7544013
158. Nair, S., Kheirkhahan, M., Davoudi, A., Rashidi, P., Wanigatunga, A. A., Corbett, D. B., et al. (2016). "ROAMM: a software infrastructure for real-time monitoring of personal health," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services, Healthcom 2016* (Munich). doi: 10.1109/HealthCom.2016.7749479
159. Oliveira, T., Leite, D., and Marreiros, G. (2016). "PersonalFit–fitness app with intelligent plan generator," in *ACM International Conference Proceeding Series* (New York, NY). doi: 10.1145/2948992.2949014
160. Ren, X., Brankaert, R., Visser, V., Offermans, S., Lu, Y., and Nagtzaam, H. (2016). "FLOW pillow: exploring sitting experience towards active ageing," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI 2016* (New York, NY), 706–713. doi: 10.1145/2957265.2961841
161. Renfree, I., and Cox, A. (2016). "Tangibly reducing sedentariness in office workers," in *Proceedings of ACM CHI 2016*. Retrieved from: https://tangibles4health.files.wordpress.com/2016/02/ian-renfree-anna-cox-tangibles-for-health-workshop-chi-2016-camera-ready.pdf
162. Skriloff, S. J. G., Gonzalez, D. C., Christensen, K. C., Bentley, L. J., and Mortensen, C. V. (2016). "FitPlay games: increasing exercise motivation through asynchronous social gaming," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (New York, NY). doi: 10.1145/2851581.2890367
163. Synnott, J., Rafferty, J., and Nugent, C. D. (2016). "Detection of workplace sedentary behavior using thermal sensors," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2016* (Orlando, FL), 5413–5416. doi: 10.1109/EMBC.2016.7591951
164. Taylor, W. C., Paxton, R. J., Shegog, R., Coan, S. P., Dubin, A., Page, T. F., et al. (2016). Impact of booster breaks and computer prompts on physical activity and sedentary behavior among desk-based workers: a cluster-randomized controlled trial. *Prev. Chron. Dis.* 13:160231. doi: 10.5888/pcd13.160231
165. Urda, J., Lynn, J., and Larouere, B. (2016). An intervention to reduce sedentary time and change perceived wellness in women office workers. *Med. Sci. Sports Exerc.* 48:595. doi: 10.1249/01.mss.0000486785.36151.36
166. Boateng, G., Batsis, J. A., Halter, R., and Kotz, D. (2017). "ActivityAware: an app for real-time daily activity level monitoring on the Amulet wrist-worn device," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017* (Kona, HI). doi: 10.1109/PERCOMW.2017.7917601
167. Cambo, S. A., Avrahami, D., and Lee, M. L. (2017). "BreakSense: combining physiological and location sensing to promote mobility during work-breaks," in *Proceedings of the 2017 Acm Sigchi Conference on Human Factors in Computing Systems (Chi'17)* (New York, NY). doi: 10.1145/3025453.3026021
168. Economou, D., Dwek, M., Roberston, C., Elliott, B., Kounenis, T., Azimi, T., et al. (2017). "PhytoCloud: a gamified mobile web application to modulate diet and physical activity of women with breast cancer," in *Proceedings–IEEE Symposium on Computer-Based Medical Systems* (Thessaloniki). doi: 10.1109/CBMS.2017.164
169. Fahim, M., Baker, T., Khattak, A. M., and Alfandi, O. (2017). "Alert me: enhancing active lifestyle via observing sedentary behavior using mobile sensing systems," in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services, Healthcom 2017* (Dalian), 1–4. doi: 10.1109/HealthCom.2017.8210838
170. Grundgeiger, T., Pichen, J., Häfner, J., Wallmann-Sperlich, B., Löffler, D., and Huber, S. (2017). "Combating sedentary behavior: an app based on a distributed prospective memory approach," in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems–CHI EA '17* (New York, NY), 1632–1639. doi: 10.1145/3027063.3053094

171. Guida, D., Basukoski, A., and Performance Database (2017). "Weightbit: an advancement in wearable technology," in *Proceedings–IEEE Symposium on Computer-Based Medical Systems* (Thessaloniki), 672–677. doi: 10.1109/CBMS.2017.85
172. Gui, X., Chen, Y., Caldeira, C., Xiao, D., and Chen, Y. (2017). "When fitness meets social networks: investigating fitness tracking and social practices on WeRun," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2017)* (New York, NY). doi: 10.1145/3025453.3025654
173. Guo, X., Liu, J., and Chen, Y. (2017). "FitCoach: virtual fitness coach empowered by wearable mobile devices," in *Proceedings–IEEE INFOCOM* (Atlanta, GA). doi: 10.1109/INFOCOM.2017.8057208
174. Hadgraft, N. T., Willenberg, L., LaMontagne, A. D., Malkoski, K., Dunstan, D. W., Healy, G. N., et al. (2017). Reducing occupational sitting: workers' perspectives on participation in a multi-component intervention. *Int. J. Behav. Nutr. Phys. Act.* 14:73. doi: 10.1186/s12966-017-0530-y
175. Klein, M. C. A., Manzoor, A., and Mollee, J. S. (2017). Active2Gether: a personalized m-health intervention to encourage physical activity. *Sensors* 17, 1–16. doi: 10.3390/s17061436
176. Wölfel, M. (2017). "Acceptance of dynamic feedback to poor sitting habits by anthropomorphic objects," in *PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY). doi: 10.1145/3154862.3154928
177. Altmeyer, M., Lessel, P., Hosseini, S., and Krueger, A. (2018a). "Silvercycling: evaluating persuasive strategies to promote physical activity among older adults," in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility–DIS '18* (New York, NY). doi: 10.1145/3197391.3205410
178. Altmeyer, M., Lessel, P., Sander, T., and Krüger, A. (2018b). Extending a gamified mobile app with a public display to encourage walking. *Psychol. Comp. Sci.* 20–29. doi: 10.1145/3275116.3275135
179. Bascur, A., Rossel, P., Herskovic, V., and Martínez-Carrasco, C. (2018). Evitapp: persuasive application for physical activity and smoking cessation. *Proceedings* 2:1208. doi: 10.3390/proceedings2191208
180. Boj, C., Díaz, D., Portalés, C., and Casas, S. (2018). Video games and outdoor physical activity for the elderly: applications of the hybrid play technology. *Appl. Sci.* 8:1912. doi: 10.3390/app8101912
181. Buckers, T., Eisemann, E., and Lukosch, S. (2018). "VRabl: stimulating physical activities through a multiplayer augmented reality sports game," in *The First Superhuman Sports Design Challenge* (New York, NY). doi: 10.1145/3210299.3210300
182. Burkow, T. M., Vognild, L. K., Johnsen, E., Bratvold, A., and Risberg, M. J. (2018). Promoting exercise training and physical activity in daily life: a feasibility study of a virtual group intervention for behaviour change in COPD. *BMC Med. Inform. Decis. Mak.* 18:136. doi: 10.1186/s12911-018-0721-8
183. Clavel, C., Whittaker, S., Blacodon, A., and Martin, J. C. (2018). "WEnner: a theoretically motivated approach for tailored coaching about physical activity," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers–UbiComp '18* (New York, NY). doi: 10.1145/3267305.3274190
184. Esakia, A., Mccrickard, D. S., Harden, S., Horning, M., Tech, V., Tech, V., et al. (2018). "FitAware: promoting group fitness awareness through glanceable smartwatches," in *ACM International Conference on Supporting Group Work* (New York, NY), 178–183. doi: 10.1145/3148330.3148343
185. Francillette, Y., Bouchard, B., Boucher, E., Gaboury, S., Bernard, P., Romain, A. J., et al. (2018). "Development of an exergame on mobile phones to increase physical activity for adults with severe mental illness," in *ACM PErvasive Technologies Related to Assistive Environments (PETRA)* (At Corfu). doi: 10.1145/3197768.3201521
186. Gouveia, R., Karapanos, E., and Hassenzahl, M. (2018). "Activity tracking *in vivo*," in *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY), doi: 10.1145/3173574.3173936
187. Gupta, A., Heng, T., Shaw, C., Li, L., and Feehan, L. (2018). "Designing pervasive technology for physical activity self-management in arthritis patients," in *The 12th EAI International Conference* (New York, NY). doi: 10.1145/3240925.3240956
188. Harrington, C. N., Wilcox, L., Rogers, W., and Connelly, K. (2018). "Designing health and fitness apps with older adults: examining the value of experience-based co-design," in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth 2018* (New York, NY). doi: 10.1145/3240925.3240929
189. Lee, D., Frey, G., Cheng, A., and Shih, P. C. (2018). "Puzzle walk: a gamified mobile app to increase physical activity in adults with autism spectrum disorder," in *2018 10th International Conference on Virtual Worlds and Games for Serious Applications, VS-Games 2018–Proceedings* (Wurzburg). doi: 10.1109/VS-Games.2018.8493439
190. Luo, Y., Lee, B., Wohn, D. Y., Rebar, A. L., Conroy, D. E., and Choe, E. K. (2018). "Time for break: understanding information workers' sedentary behavior through a break prompting system," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems–CHI '18* (New York, NY), 1–14. doi: 10.1145/3173574.3173701
191. Marcu, G., Misra, A., Caro, K., Plank, M., Leader, A., and Barsevick, A. (2018). "Bounce: designing a physical activity intervention for breast cancer survivors," in *PervasiveHealth '18: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY). doi: 10.1145/3240925.3240927
192. Meyer, J., Beck, E., von Holdt, K., Gansefort, D., Brand, T., Zeeb, H., et al. (2018). "ActiStairs: design and acceptance of a technology-based intervention to advocate stair-climbing in public spaces," in *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care* (New York, NY), 59–66. doi: 10.1145/3264996.3264997
193. Schäfer, H., Bachner, J., Pretscher, S., Groh, G., and Demetriou, Y. (2018). "Study on motivating physical activity in children with personalized gamified feedback," in *UMAP '18: Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (New York, NY). doi: 10.1145/3213586.3225227
194. Tajadura-Jiménez, A., Cuadrado, F., Rick, P., Bianchi-Berthouze, N., Singh, A., Väljamäe, A., et al. (2018). "Designing a gesture-sound wearable system to motivate physical activity by altering body perception," in *MOCO '18: Proceedings of the 5th International Conference on Movement and Computing* (New York, NY). doi: 10.1145/3212721.3212877
195. van Dantzig, S., Bulut, M., Krans, M., van der Lans, A., and de Ruyter, B. (2018). "Enhancing physical activity through context-aware coaching," in *PervasiveHealth '18: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*. doi: 10.1145/3240925.3240928
196. Ciravegna, F., Gao, J., Ireson, N., Copeland, R., Walsh, J., and Lanfranchi, V. (2019). "Active 10–brisk walking to support regular physical activity," in *Proceedings Paper in Pervasive Health* (New York, NY). doi: 10.1145/3329189.3329208
197. Geurts, E., Van Geel, F., Feys, P., and Coninx, K. (2019). "WalkWithMe–personalized goal setting and coaching for walking in people with multiple sclerosis," in *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization–UMAP '19* (New York, NY), 51–60. doi: 10.1145/3320435.3320459
198. Martin-Niedecken, A. L. (2019). "Plunder planet- an adaptive fitness game setup for children," in *Extended Abstracts of the 2019 CHI Conference* (New York, NY), 19–30. doi: 10.1145/3290607.3311765
199. Oyibo, K. A., Olagunju, H., Olabenjo, B., Adaji, I., Deters, R., and Vassileva, J. (2019). "BEN'FIT: design, implementation and evaluation of a culture-tailored fitness app," in *Adjunct Publication of the 27th Conference* (New York, NY). doi: 10.1145/3314183.3323854
200. Samariya, A., Sharma, A., Fitriani, M., Ferguson, T., Fails, J. A., and Alan, J. (2019). "KidLED: a colorful approach to children's activity awareness," in *IDC '19: Proceedings of the 18th ACM International Conference on Interaction Design and Children* (New York, NY). doi: 10.1145/3311927.3326594
201. van Kasteren, Y., Champion, S., Perimal-Lewis, L., and Sartor, J. B. (2019). "Thermal comfort and physical activity in an office setting," in *ACSW 2019: Proceedings of the Australasian Computer Science Week Multiconference* (New York, NY). doi: 10.1145/3290688.3290733

202. Wang, Y., and Reiterer, H. (2019). "The point-of-choice prompt or the always-on progress bar? A pilot study of reminders for prolonged sedentary behavior change," in *CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY). doi: 10.1145/3290607.3313050
203. Zhang, J., and Jemmott, J. B. III. (2019). Mobile app-based small-group physical activity intervention for young African American women: a pilot randomized controlled trial. *Prev. Sci.* 20, 863–872. doi: 10.1007/s11121-019-01006-4
204. Orji, R., Lomotey, R., Oyibo, K., Orji, F., Blustein, J., and Shahid, S. (2018). Tracking feels oppressive and 'punishy': exploring the costs and benefits of self-monitoring for health and wellness. *Digit. Health* 4:2055207618797554. doi: 10.1177/2055207618797554
205. Fogg, B. J. (2009). "A behavior model for persuasive design," in *Proceedings of the 4th International Conference on Persuasive Technology–Persuasive '09* (New York, NY). doi: 10.1145/1541948.1541999
206. Glanz, K., and Rimer, B. K. (2005). *Theory at a Glance: A Guide for Health Promotion Practice*. U.S. Department of Health and Human Services, National Institutes of Health.
207. Gerling, K. M., Livingston, I. J., Nacke, L. E., and Mandryk, R. L. (2012). "Full-body motion-based game interaction for older adults," in *Conference on Human Factors in Computing Systems–Proceedings* (New York, NY). doi: 10.1145/2207676.2208324
208. Orji, R., Mandryk, R. L., Vassileva, J., and Gerling, K. M. (2013). "Tailoring persuasive health games to gamer type," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems–CHI '13* (New York, NY). doi: 10.1145/2470654.2481341
209. Orji, R., Mandryk, R. L., and Vassileva, J. (2017). Improving the efficacy of games for change using personalization models. *ACM Trans. Comput. Hum. Interact.* 24:32. doi: 10.1145/3119929
210. Purpura, S., Schwanda, V., Williams, K., Stubler, W., and Sengers, P. (2011). "Fit4life: tje design of a persuasive technology promoting healthy behavior and ideal weight," in *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems–CHI '11* (New York, NY). doi: 10.1145/1978942.1979003
211. Vlieghe, J. and De Troyer, O. (2020). *Report D3: Literature Study on Persuasive Techniques and Technolog*. Retrieved from: https://wise.vub.ac.be/tickle/wp-content/uploads/2015/12/Report-D3_final.pdf

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 02 July 2020
doi: 10.3389/frai.2020.00045

Opinion Formation on the Internet: The Influence of Personality, Network Structure, and Content on Sharing Messages Online

Laura Burbach, Patrick Halbach, Martina Ziefle and André Calero Valdez*

Human Computer Interaction Center, RWTH Aachen University, Aachen, Germany

Today the majority of people uses online social networks not only to stay in contact with friends, but also to find information about relevant topics, or to spread information. While a lot of research has been conducted into opinion formation, only little is known about which factors influence whether a user of online social networks disseminates information or not. To answer this question, we created an agent-based model and simulated message spreading in social networks using a latent-process model. In our model, we varied four different content types, six different network types, and we varied between a model that includes a personality model for its agents and one that did not. We found that the network type has only a weak influence on the distribution of content, whereas the message type has a clear influence on how many users receive a message. Using a personality model helped achieved more realistic outcomes.

Keywords: opinion formation, personality traits, message spread, social networks, network types, latent process model

OPEN ACCESS

Edited by:
Panagiotis Germanakos, SAP SE, Germany

Reviewed by:
Yilun Shang, Northumbria University, United Kingdom
Mattia Zanella, University of Pavia, Italy

****Correspondence:***
Laura Burbach
burbach@comm.rwth-aachen.de

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: *31 October 2019*
Accepted: *25 May 2020*
Published: *02 July 2020*

Citation:
Burbach L, Halbach P, Ziefle M and Calero Valdez A (2020) Opinion Formation on the Internet: The Influence of Personality, Network Structure, and Content on Sharing Messages Online. Front. Artif. Intell. 3:45. doi: 10.3389/frai.2020.00045

1. INTRODUCTION

Social networks such as Facebook, Instagram and Twitter are now integrated into most people's everyday lives. The users of social networks no longer just use them to keep in touch with friends, but increasingly facilitate social networks to search for information. Users also form opinions based on the information and contributions available in social networks. While searching for information and integrating it into their opinion formation, users are no longer just passive recipients of information in online social networks, but are also actively spreading their own opinions (Hóllig and Hasebrink, 2016; Li et al., 2017; Frees and Koch, 2018). Thus, the dissemination of new information is increasing and a broader range of opinions is voiced (Bakshy et al., 2012).

Social networks play a powerful role and not only influence the formation of opinion of individuals, but can also play a decisive role in political situations and decisions (Guille et al., 2013). It has been shown, that social networks have a strong influence on political decisions. One example for this was the American presidential election in 2008, where many people perceived a strong influence of Twitter on the elections (Hughes and Palen, 2009; Shang, 2019).

While the amount of information users receive has changed through social networks, we also have to consider that information can now be personalized through the individual users' interaction with the network and its structure (DeVito, 2017). On the Internet, users can find almost any information they are looking for. However, the amount of information available on the Internet is now so large that users are no longer able to consume all the information. In addition, users also find contradicting information on the Internet. The increasing availability of information on the Internet has led to the development of recommendation systems (Adomavicius and Tuzhilin, 2005). Aiming to make it easier for users to select information, these systems analyse the information available, filter it according to specific criteria and provide users with recommendations tailored to their needs (Burke, 2002).

While this makes the search for information on the Internet and in social media easier for users at first, this selection of information can also have negative consequences. In the past, for example, the number of voices that view social networks as something negative has increased due to the fact that political opinions have been deliberately influenced and political results manipulated (Stark et al., 2017; Shang, 2019).

So far, research does not tell us how much opinion-forming processes actually take place in social networks and we can not predict those processes, let alone their consequences, yet. However, it is very important to consider the impact that the use of social networks has on user opinion formation.

Opinion forming and the processes that influence the formation of opinions have been studied since the 1960s. In the meantime, aspects that were of little or no importance before the rise of the Internet have become relevant for the research of the formation of opinion by individuals. First of all, in social networks every user can express his opinion and reach a large number of other users by just resharing or reposting (Cheng et al., 2014). By other users reposting the opinion of a user, long cascades are created and the opinion is disseminated (Cheng et al., 2014). Users share information targeting to convince other users of their opinion. Social media simplified that users can convey one's opinion to other individuals.

Still, users need to be connected with the users that they want to convince. Each user's network is individual and different from the network of other users, and they themselves can expand or shrink the network, which in turn can affect the accessibility of users.

Nevertheless, one aspect that needs to be considered is that users differ in their sharing habits. Lottridge and Bentley (2018) investigated the motivation to share contents and the frequency of sharing news on public, social and private platforms. They differentiate between different types of forwarding. Users can share messages with other users, they can share information in a personal message or share information on a social network or public a content publicly. In their research, they found that users have different intentions with different forms of sharing. Users share information publicly, primarily when they want to contribute an ideology. In contrast, they send private messages primarily to tell stories that correspond to their own interests or the context in which the user finds themselves. The first group shared news in all channels; they share both publicly, socially, and privately. In contrast, the second group does not share news at all. The last group shares messages only in private and social channel. Matching this, they also found that the group that shared the least posts had a negative attitude toward online discussion, whereas the group that shared the most posts had a neutral attitude toward online discussion (Lottridge and Bentley, 2018).

To understand how the dissemination of information and thus also the formation of opinion in social networks takes place, it is necessary to first consider what motivates users have to express their opinion in social media and which personality traits the users have that publish their opinion in social media. It is further relevant to look at different network structures, as they can also influence how content is spread and to whom—to friends or other users.

2. RELATED WORK

In this study, we consider what influences how messages are spread in online social networks using an agent-based simulation. Therefore, we explain what is known in theory about the spread of information and the formation of opinions in this section. We further introduce the latent process model, on which we built our simulation and explain further aspects that are important for the agent-based simulation.

2.1. The Study of Complex Systems

The consideration of the spread of news in social networks is based on a complexity. We can also speak of a complex social system. In other words, a system consisting of several ontological levels. This system can be divided into its micro- and macro-level, which represent interacting subsystems (Conte et al., 2012). To understand complex social systems, it is not enough to look at the individual parts and understand them, but the overall system is more than the sum of all individual parts. If a system or a behavior cannot be described by the individual parts or subsystems alone, but in the overall system more becomes visible, one also speaks of *emergence* or emergent behavior.

A helpful way to understand this emergent behavior is to simulate the individual subsystems (Epstein, 2007). In this study, we also simulate the subsystems or processes of the spread of information in online social networks to become an understanding of the overall system. To do so, we use agent-based modeling, what is a well-suited method here (Epstein, 2007; Calero Valdez and Ziefle, 2018). One advantage of the system-theoretical approach with agent-based models is that we can use it to simulate how networks are created and information is disseminated or to simulate similar processes. Rational choice models often play a role in this type of modeling (Gilbert, 2008). We also developed such a model for this study. Agent-based models are not created aiming at an exact representation of the real world, but they try to represent individual behavior as realistically as possible and thus always simplify reality. The models also enable a qualitative observation of the behavior of the system. Evaluating such models is difficult and requires an independent replication of the model as well as a comparison with other models and a validation (Rouchier et al., 2008).

2.2. Information and Opinions

First, we must notice, that whether or not information is spread in a social network using the technological infrastructure, is independent from the spread of an opinion in the users minds. Therefore, it makes sense to model both sides of this process, first information dissemination and second opinion formation.

2.3. Spread of Information

Research speaks of information dissemination, information spread, or information diffusion when a person or a group of people sends information in a network (Li et al., 2017). Information dissemination has already been analyzed in many ways and it has been considered which aspects influence information processing as well as which information is processed how fast and in what manner (Christakis, 2007; Zhang and Wu, 2011). There are a number of dissemination models and other methods that are used to understand the diffusion phenomenon.

The spread of information in networks is similar to the spread of disease in contact networks, however, while the latter requires a face-to-face interaction—thus have relatively low limit on the edge-degree of nodes—the former can spread much faster due to the fact that online social networks allow for thousands of followers. When the president of the United States retweets a post from a user, several million other users are immediately exposed to this type of information.

When social networks and the structure of social networks are analyzed, it is also possible to examine the relationship between individual users and to identify patterns in user interactions (Wasserman and Faust, 1994). Some studies, for example, have been concerned with finding opinion leaders. Java et al. (2007) have shown how influential bloggers can be identified and Goyal et al. (2008) have identified opinion leaders in social networks through actions and interactions.

While much about the structure of networks on opinion formation has already been studied mathematically (Albi et al., 2016; Toscani et al., 2018), little is known about the psychological reasons for the spread of information in online social networks. For example, we do not know why the information in social networks flows in a certain direction. In addition, while we know that opinion leaders exist, we know little about how much influence they have on opinion formation and who are the most important users in disseminating information apart from the opinion leaders. Also only little is known about which factors influence the information diffusion process (Li et al., 2017).

2.3.1. Diffusion Models

So far, diffusion models are used for many different purposes. A use case that is relevant for us is how messages are spread or how the spread of messages can be stopped (Guille et al., 2013). Following, we explain some basics of diffusion models. Basically, the *diffusion process* can be divided into two basic components. The first basic component of the process is a certain structure. The structure consists of a diffusion graph. The graph shows who influences whom. The second basic component is the temporal dynamics of the *diffusion process*. It describes how the diffusion rate develops. The diffusion rate means how many nodes take over the information over time. In the course of the *diffusion process*, a node can either be activated or not. An activated node has received the information and is trying to spread it. Within a network a successive activation of nodes takes place, which is called *diffusion process*. In models that consider the dissemination of information on social networks, users are usually influenced only by the people they are connected to. It is assumed that information is disseminated through information cascades (Guille et al., 2013).

2.4. Opinions and Attitudes

Since the spread of information does not equate the spread of opinions, we must understand how opinions are formed. Opinions are typically voiced—they are public. However, opinions may differ from the attitude of a person. The internal attitude may differ from the external opinion.

Moreover, the term attitude refers to various phenomena. There is no uniform understanding of the term attitude, let alone a uniform definition. There is also no agreement as to whether the terms opinion and attitude are synonymous or different. There is the opinion that both terms mean the same and are interchangeable, but also the opinion that they describe related processes, but refer to different aspects of these processes (Meinefeld, 1977; Oskamp and Schultz, 2005). However, there is agreement that attitude is a tendency to evaluate an object positively or negatively and to react to it if necessary. In this article, we are also guided by this notion of attitude, which also corresponds to the definition of Oskamp and Schultz.

We used the latent process model by DeFleur and Westie (1963) as the theoretical basis for the simulation. It explains the emergence of attitudes. According to this model, attitude is a theoretical construct, whose state should be considered as unknown. DeFleur and Westie see the attitude as a process variable within the opinion forming process. The opinion forming process is a preceding process to the reaction as a following process. The three processes form together the latent attitude. The latent attitude of a person is visible due to an observable reaction of the person. The reaction can be cognitive, a change in belief, affective, a change in emotion, or behavioral, a change in interaction.

For example, reading a post about the Iranian missile launch on American Forces in Iraq on January 8th 2020, claiming that no damage was done to American soldiers, could lead to several reactions. The reader could change their belief about the severity of the conflict situation between USA and Iran; they could perceive an emotional relief about the severity, yet cognitively perceive the threat as equally strong; or they could post contradictory or agreeing information online. In the latent process model, the affective and cognitive component are relatively independent of one another. The behavioral component is governed by both cognitive and affective processes.

2.4.1. Three-Component Model of Attitude

Attitudes can consist of a cognitive, affective and behavioral component (Oskamp and Schultz, 2005). The cognitive component is a person's thought of an adjustment object. This component is also called conviction. While the cognitive component refers to what a person thinks, the affective component relates to what a person feels. More precisely, the component incorporates the feelings or emotions toward an object. Ultimately, the behavioral component holds the concrete intentions of a person and how the person actually behaves toward an object (Eagly and Chaiken, 1993; Oskamp and Schultz, 2005; Hartung, 2006). If the three different components are compared, the importance of the affective component in particular can be emphasized because emotions are motivating and make a person behave more strongly than cognition (Oskamp and Schultz, 2005).

2.4.2. The Consistency Theorem

The processes described above are also referred to as the *three-component model* (McGuire, 1985; Eagly and Chaiken, 1993). In the past, however, some criticism has also been voiced about this established model. One of them stresses that it is not clear how the individual processes or components relate to each other. Thus, there is disagreement as to whether the three components actually say the same thing or whether, contrary to this opinion, they differ so much from each other that it is better to divide them into three separate units. In addition, there is the opinion that the attitude does not always consist of all three components (Oskamp and Schultz, 2005). Meinefeld (1977) and Oskamp and Schultz (2005) assume that the three components are just different names for the same thing are considered proponents of the *Consistency Theorem*. On the other hand, the proponents of the *Separate Entities Model* reject the Consistency Theorem and consider the three components as separate processes (Fishbein and Ajzen, 1975).

2.4.3. The Latent-Process Model

The basis for the *latent process model* was the emergence of the critique of the *Consistency Theorem* and the *Separate Entities Model*. The three models are connected by the fact that they all try to explain how attitudes arise. While the *Consistency Theorem* assumes a consistency between attitude and behavior, DeFleur and Westie criticize exactly this assumption of consistency. In their opinion, an inner process takes place between the appearance of an external stimulus and a behavior. This inner process cannot be observed directly (through visible behavior), but the *Consistency Theorem* and the *Separate Entities Model* aim to explain this inner process.

On the contrary, DeFleur and Westie are of the opinion that the visible behavior is not the same as this inner process. They also assume that in addition to attitude other (social) factors influence behavior. In their opinion, the theoretical construct "attitude" is a link to describe the connection between object and behavior. The attitude itself must, however, be regarded as unknown. In the *latent process model* there is also an unobservable process of attitude formation which takes place before the inner process or attitude. Furthermore, there is a reaction that follows the attitude. In her opinion, a stimulus triggers cognitive and affective processes and also the process of behavioral intention. Then either individual processes or a combination of the processes form the latent attitude. This latent attitude becomes visible through a cognitive or affective reaction or behavior. In this model, the attitude can therefore be regarded as a probability conception and says how likely a person is to behave toward an object similar to how he has behaved in the past. As a result, in the model, attitude does not necessarily explain how a person behaves, but rather shows the regularity of certain behavior patterns. In addition to this advantage of the model, another advantage is that no relationship between the individual processes is assumed. In contrast, it is possible that only one process takes place, but also that two or all three processes take place (DeFleur and Westie, 1963; Oskamp and Schultz, 2005).

Other Researchers (e.g., Xiong and Liu, 2014) have investigated opinion formation using latent internal opinions. However, to our knowledge none have investigated the differences in processes underlying the disparity between opinion and behavior using a process model.

2.5. Social Networks and Their Users

To consider how messages are spread in a social network, not only the nature of the network must be considered, but also the personality of the users. It makes sense to model the users and their personality as realistically as possible to be able to make replicable statements. Some studies (Bachrach et al., 2012; Kosinski et al., 2013; Dong et al., 2014) in the past have already pointed out that the personality of users of online social networks is related to the characteristics of the respective network. Once the users have been created as realistically as possible, the simulation can start. This is followed by an arbitrary or fixed number of simulation steps. In the individual steps of the simulation, they interact with other users and with the environment in which the users live. Some parameters are set to determine how likely which stochastic processes occur (Serrano and Iglesias, 2016).

We consider personality traits of users of online social networks and how users behave in social networks as a basis for the most truthful possible design of agents in our model. For this study, we use the Big Five personality model to design our social network users, because it is the most established model to describe the personality of individuals. Following, we first describe the Big Five personality model and then how these personality traits are correlated with the use of or behavior in online social networks.

2.5.1. Big Five Personality

There are many different models that try to describe the personality of individuals. If one wants to describe the personality of individuals, one inevitably comes across the Big Five personality trait model. It is a very established concept to describe different personalities (Costa and Mccrae, 1992). The personality of the individual is described in the model on the basis of five characteristics: *Openness to experience*, *conscientiousness*, *extraversion*, *agreeableness* and *neuroticism*. *Openness* means that a person has a lot of imagination and intellectual curiosity. *Conscientiousness* is understood to mean that a person is careful and well-organized. A person with a strong *extraversion* personality trait is sociable and tends to look for simulation. The personality trait *neuroticism* refers to negative emotions such as anxiety and depression and is defined as emotional instability. An individual, with a pronounced *agreeableness*, is very cooperative and has a lot of compassion for his other people (Power and Pluess, 2015).

Although the *Big Five factors* were initially designed as individual personality traits, some studies have shown that the

TABLE 1 | Big Five Personality traits and relations to each other.

Personality trait	1	2	3	4
Extraversion	–			
Agreeableness	**0.35*** **	–		
Conscientiousness	**0.15*** **	**0.27*** **	–	
Neuroticism	**−0.24*** **	**−0.05*** **	**−0.20*** **	–
Openness	**0.41*** **	**0.22*** **	**0.24*** **	**−0.09*** **

*Taken from Power and Pluess (2015). Statistically significant estimates are in bold. *p < 0.05, **p < 0.01, and ***p < 0.001.*

traits are interrelated (Watson and Humrichouse, 2006; Grant and Langan-Fox, 2007). Power and Pluess investigated 5,011 European adults in their study and for the first time investigated the common heredity of the five *Big Five personality traits* with a GREML (Genomic-relatedness—Matrix Residual Maximum Likelihood) approach (Power and Pluess, 2015). They found that all the personality traits correlate with each other and also that all of them, except of *openness* correlate with gender.

Of course, there are other models that describe the personality of individuals, but in this study we focus on the model of the Big Five personality traits as the most established model to describe personality.

2.5.2. Personality and Social Networks

In 2011, two studies by Gosling et al. looked at the use of Facebook and also measured how the *Big Five personality traits* of users related to their use. In the first study, the participants gave self-disclosure about their Facebook use. The study showed, that users with high *extraversion* values are connected to many Facebook friends. In addition, *extroverted* persons comment more frequently on contributions. While a high level of *conscientiousness* is associated with people spending less time on Facebook, more *open-minded* people tend to post more photos on Facebook compared to other users (Gosling et al., 2011).

In the second study, the Facebook profiles of the respondents were quantified in advance by "observers" and evaluated with regard to possible personality traits to reduce the effect of the self-report. While they did not find the results on *conscientiousness* in this study, they again found the same results on *openness* and *extraversion* (Gosling et al., 2011). Confirming the results of Gosling et al., other studies tried to even forecast the personality of users based on their facebook profiles (Golbeck et al., 2011)

Bachrach et al. conducted a study in 2012 with 180,000 participants and thus a significantly larger sample. They found similar results as in the previously described studies. For example, they found that more *open* people also publish and like posts more frequently and join Facebook groups more often. In addition, they found out that people who are more *conscientious* mark less post with like, but publish many photos. As with the studies described above, further studies showed, that extroverted individuals are associated with more Facebook friends, publish and like posts more frequently (Cullen and Morse, 2011; Bachrach et al., 2012; Cheevasuntorn et al., 2017).

2.6. Modeling a Social Network

In addition to the agents of a simulation, the environment must also be simulated. In the environment—the area in which the agents "live"—the agents interact with each other (Serrano and Iglesias, 2016). For modeling social networks in simulation environments the structure of the network has to be mirrored into an artificial environment. This can be done by either replicating a real social network or by referring to artificial network topologies that has similar characteristics to real social networks.

The generation of artificial networks has been investigated since the 1960s (Wasserman and Pattison, 1996). Usually these models are based on real social networks (Leskovec et al., 2009). One difficulty in depicting online social networks is that they are usually large and have both insecure structures and overlapping groups. However, this difficulty can be overcome with the help of agent-based modeling. With this method, networks with similar properties can be built generatively (Barrett et al., 2009; Pham et al., 2013). Agent-based models also enable to examine a large number of networks with similar characteristics and thus to simulate real network behavior.

An important basis to model message spreading is the structure of the network. The real social affiliation can also be seen in the structures of social online networks (Zheleva et al., 2009). Therefore, we also used different social network structures to connect our agents with each other in the environment, i.e., the social network.

2.7. Network Topologies

Network topologies serve as a structural basis of social networks as they make it possible to understand the formation of node and link distribution and to describe effects that occur depending on the structure (see **Figure 1**). Network topologies can be classified into three types of networks: (1) Random Graph, (2) Scale-free, and (3) Small-World Networks that follow their own particularities (Albert and Barabási, 2002). In general it is important to know that several measurements can be provided to describe a network such as its centrality, cluster coefficient, and average path length. The clustering coefficient is an important value for examining the extent to which a network consists of local, strongly interconnected groups.

2.7.1. Random Graph Networks

A random graph network is created by starting with a fixed set of vertices and adding edges between those vertices randomly. The most popular random graph model is the Erdős–Rényi model and actually combines two closely related models. The first model was proposed by Paul Erdős and Rényi and makes all graphs on fixed sets of vertices with a fixed number of edges equally likely. Contrasting, the second model proposed by Gilbert provides a fixed probability for each edge to exist or not, independently of the other edges (Erdős and Rényi, 1960).

2.7.2. Small-World Networks

In a Small-World network model most nodes are indirectly adjacent to each other. This means that the average path length between the nodes is rather small, as every path

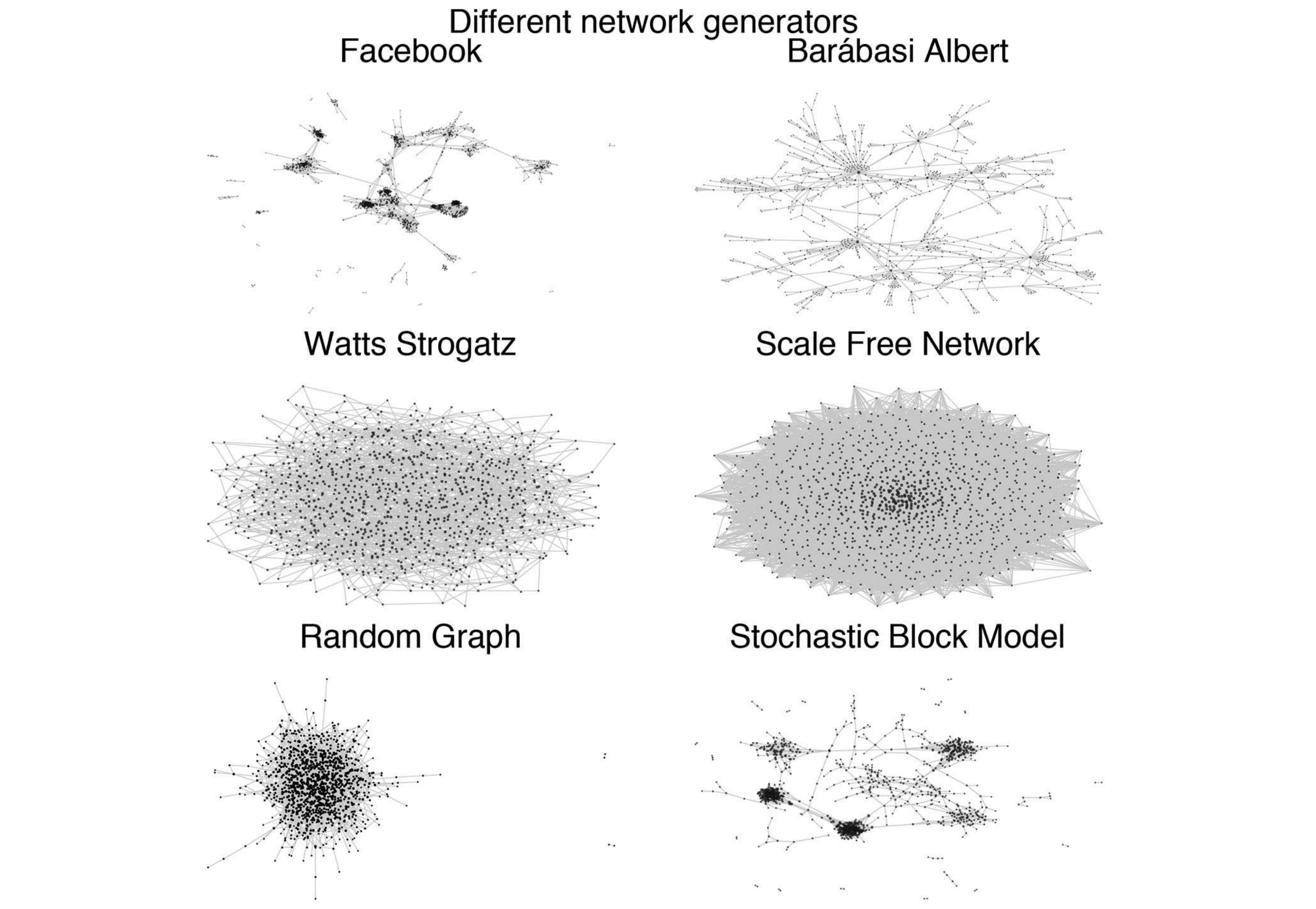


FIGURE 1 | Different sample network topologies for 1,000 vertices: (top left) real Facebook data, (top right) Barábasi-Albert model, (bottom left) Watts Strogatz mode, (bottom right) Scale-free network. All Networks are displayed using a force-based layout.

between two nodes requires only a small number of hops. This implementation is needed to realize the small-world phenomenon that for example was investigated by Milgram and is connected to the idea that several networks follow the rule of "six degrees of separation" (Milgram, 1967). Small-World network properties appear in many real-world networks (Watts and Strogatz, 1998).

The Watts-Strogatz Model addresses this approach by extending the Erdős–Rényi model with an algorithm to create local clustering and triadic closure. It constructs a network with a regular ring lattice and rewires in a next step the vertices with a probability β for each edge while avoiding self-loops.This results in a graph with high local clustering and compared to a regular random network significantly reduced average path lengths through randomly rewired links. A minor drawback of this network is its weakness in producing realistic degree distributions as no hubs or scale-free distributions can be created.

2.7.3. Scale-Free Networks

So-called scale free networks follow a power law distribution in terms of the network degree of their nodes. This means, that the degree of a node is proportionally related to its probability to get new connections. This results in an 80-20 distribution for the degree of nodes as 20% of the nodes are dominating 80% of the other nodes concerning their degree. This topology is predominant in social networks as for example the Erdős number shows: it describes, how close two scientists are in terms of collaboration, measured through their other collaborators in publications (Newman, 2001).

Such a network requires two main features considering its evolution process: it has to grow over time and the addition of nodes has to follow a preferential attachment strategy so that the probability for connecting a new node to the existing ones is higher for nodes who already have a high node degree than for those with low degree.

The Barábasi-Albert model describes a typical scale-free network topology. The algorithm serves for a power-law distribution of node degrees, resulting in a little amount of very well-connected hubs and a majority of nodes with only few connections to other nodes.

2.8. Stochastic Block Model

The Stochastic Block Model (Mossel et al., 2012) assumes that not all nodes stem from the same class. Speaking in network terms, users are from different countries, cities, social identity groups. The probability of nodes being attached to other nodes may now differ depending on class. Typically, nodes from the same class are highly interconnected, while only few connections are formed between classes. Using a stochastic block model community structures in the data can be explained, as found in real networks such as Facebook.

2.9. Facebook

Facebook is designed to allow users to maintain their connections, share content, and interact with content. However, Facebook was not intended to deliver news (Facebook, 2012). Regardless of what Facebook's original goal was and how the news feed was designed to serve that goal, the news feed is playing an increasingly important role in users' information flows. This is due to the fact that the user base has increased dramatically and users are increasingly integrating the news feed as part of their everyday lives (Duggan et al., 2014). Facebook is also becoming increasingly important as a source of news information. By 2014, 41 percent of Americans were already consuming news on Facebook (Matsa and Mitchell, 2014; DeVito, 2017).

Facebook in the role of the news source is responsible for some influential gatekeeping and agenda-setting functions that used to be more in the hands of human editors (McCombs and Shaw, 1972; DeVito, 2017). Here, Facebook intervenes primarily in the process of disseminating information by selecting which stories or topics are presented. Thus, it is no longer the editors who select the content to be presented, but the algorithms that are responsible for the selection of a story (boyd and Ellison, 2007; DeVito, 2017).

2.9.1. Facebook Network

The Stanford Social Network Project provides a dataset for a Facebook Network consisting of 4,039 nodes and 88,234 edges. This data makes it possible to obtain a non-algorithmic, realistically grown network (McAuley and Leskovec, 2012). Stanford researches used this dataset to identify different types of social circles in online social networks such as friends or family members. We include this network dataset to compare the effects occurring in the other topologies to a realistic network.

2.10. Research Aim

With this study we aim to understand, what increases or decreases the spread of messages in social networks. To look at this question, we designed an agent-based model. Using this model, we focused on three different aspects and wanted to find out, how the three aspects influence the agents willingness to share a message. As a first factor, we considered four different types of content. Secondly, we considered five different network types and lastly we either considered the personality of the agents or did not.

2.10.1. Other Types of Random Graphs

Many other types of random graphs, such as multi-type(Shang, 2016), bi-partite graphs, or stochastic block models exist. Some of these graphs might even be more suitable for the simulation of social networks. However, many of the properties about large scale components and their connectedness are similar to simple models anyways (Kang et al., 2015). As a first step, we choose to investigate single, unitype graph models in this paper.

3. METHOD

To study the effects of a dual-process model in different network settings we created an agent-based model to simulate message sending in networks using the Julia language (Bezanson et al., 2017). The simulation is written completely in Julia and available in a public GitHub Repository. Similarly, the data analysis is written in R using R Markdown and also openly available on GitHub.

3.1. Simulating Message Sending in Networks

To simulate how messages are sent in a network, we need to find ways to artificially instantiate the components that play a role in such a process. In our case we must simulate the individuals (the agents), the network, and the messages. We simplify our model, by assuming that only one message exists at a time. By running multiple simulations we can investigate the effect of different messages.

3.1.1. The Message Model

We use a very simplified type of a message model. Messages contain two values, of which one describes the affective stimulus and the other marks the cognitive stimulus of a message. Both are drawn from four different options, which can be represented as a tuple [$val = (\text{affective, cognitive})$]. We have chosen four message types that represent different the affective and cognitive values in different forms. We considered one message, that is mostly affective [*affective content*, $val = (0.8, 0.2)$], one message, that is mostly cognitive [*cognitive content*, $val = (0.2, 0.8)$], one content, that is both, affective and cognitive [*both*, $val = (0.8, 0.8)$] and lastly one weak content, that is rather neutral [*weak content*, $val = (0.2, 0.2)$].

3.1.2. The Agent Model

The core idea of our study was to investigate the effect of the dual-process model in message sending. Thus our agents have virtual representations of the dual process model.

First, agents remember their affective and cognitive attitude toward a message. These attitudes are both drawn from the uniform distribution between 0 and 1 [$U(0, 1)$]. They are assumed to be statistically independent, which is reasonable as people may have different attitudes toward a subject on a cognitive or affective level.

Second, the behavior of the agents follows two individual thresholds. The noticing threshold (drawn from $U(0, 1)$) determines how much affective stimulation it requires to notice the content. By comparing the noticing threshold with the affective value of the message, it is determined whether the message is plainly ignored or evaluated further.

These variables were drawn from a uniform distribution to simplify the opinion space to a domain of [0;1]. In another experiment, we tested using normally distributed data $N(0, 1)$, yielding very similar results. To match these variables to a domain of [0;1] we used a *arctanh*(x)-transformation.

Next, in case they noticed it, agents evaluate whether or not to forward a message. The posting threshold is compared with the full dual-process evaluation following a tripartite approach. The affective value of the message and the affective attitude both form the affective process variable by taking their mean. This entails that it requires both parts (attitude and message activation) for the process to play a role. Although, a weak activation can already trigger a relatively strong response (because we take the mean of both). The same is done for the cognitive process, i.e., take the mean of cognitive attitude and cognitive value of the message. Both processes are then combined using the geometric mean, making it necessary to have a strong activation on both processes to start the behavioral process. This last process is compared against the posting threshold. If the process value is higher than the threshold, the user forwards the message.

Simulating this process allows giving agents internal attributes and opinions that are not acted on unless a message activates them. But how to pick the thresholds?

We differentiate between two different agent models. The *random* agent, simply draws these thresholds from a uniform distribution [$U(0, 1)$]. This sets the expected threshold to 0.5 with strong variation in the sample [$E(SD) \approx 0.289$].

The other agent we call *personality* agent, because we base this agent on a personality model. The underlying model is the Big Five personality model, from which we use three dimensions (Goldberg, 1990). Agents can vary with regard to extraversion, openness, and conscientiousness. There is evidence (see section 2) toward these measures that indicates that they influence behavior in social networks.

For example, higher extraversion of an individual increases its likelihood to have many connections on Facebook (Lönnqvist and Itkonen, 2014). Further, a higher openness makes it more likely for users to notice new content on social media (Alan and Kabadayı, 2016). Lastly, a higher conscientiousness decreases the likelihood to post content without thoughtful consideration (Gumelar et al., 2018/07). Thus, we derive the following: The *noticing threshold* in the *personality* agents is the mean of a random value from $U(0, 1)$ and the inverse of the openness of the agent. This leads to a lower threshold for more open agents. The *posting threshold* is the mean of a random value from $U(0, 1)$ and the conscientiousness of the agent. This increases the threshold for more conscientious users. The extraversion of an agent is later used in the network formation.

To create realistic personality values we draw these values from a multivariate normal distribution that is generated using the correlation (**Table 1**). To ensure that our values are in the domain of [0; 1] we take the tangens hyperbolicus [which yields a domain of $(-1; 1)$], add one, and divide by 2.

This *personality* agent should behave more realistically than the completely *random* agent.

3.1.3. Six Different Network Types

To measure the effect different random networks have on message spreading we generate six different network types. For this purpose we use the respective network generators supplied in the LightGraphs package (Bromberger et al., 2017) and the SNAP Dataset (Leskovec and Krevl, 2014). The scale free network uses the LightsGraphs implementation by Cho et al. (2009), therefore also referred to as "Cho." The stochastic block model assumed 20 communities and randomly divides the agents to these communities. This is achieved by choosing 20 random numbers from a uniform distribution and dividing the by the sum. These numbers are then multiplied with the intended agent size. The weight matrix for the generator is randomly created by limiting the diagonal entries to values between 0.01 and 0.05 times the clustersize, and all other entries between 0.0001 and 0.01. For 60% of the non-diagonal entry we randomly also select 0, to achieve non-connected components. This achieves similar community structures as the facebook data used here as well.

Overview of network properties by configuration

In total 12.000 simulations were performed using 10 different settings.

Network type	# of simulations	Edges by network size (and standard error)		
		1,000 agents	2,000 agents	4,039 agents
Personality based agents				
Barabasi Albert	1,000	999 ± 0.00	$1{,}999 \pm 0.00$	$4{,}038 \pm 0.00$
Facebook	1,000	$5{,}409 \pm 11.16$	$21{,}629 \pm 26.24$	$88{,}234 \pm 0.00$
Random	1,000	$1{,}996 \pm 0.06$	$3{,}996 \pm 0.06$	$8{,}074 \pm 0.06$
Stochastic block model	1,000	$1{,}941 \pm 11.89$	$3{,}911 \pm 17.68$	$8{,}033 \pm 25.52$
Scale Free (Cho et al., 2009)	1,000	$2{,}000 \pm 0.00$	$4{,}000 \pm 0.00$	$8{,}078 \pm 0.00$
Watts Strogatz	1,000	$2{,}000 \pm 0.00$	$4{,}000 \pm 0.00$	$8{,}078 \pm 0.00$
Random agents				
Barabasi Albert	1,000	999 ± 0.00	$1{,}999 \pm 0.00$	$4{,}038 \pm 0.00$
Facebook	1,000	$5{,}397 \pm 11.16$	$21{,}608 \pm 27.06$	$88{,}234 \pm 0.00$
Random	1,000	$1{,}996 \pm 0.06$	$3{,}996 \pm 0.07$	$8{,}074 \pm 0.06$
Stochastic block model	1,000	$1{,}934 \pm 12.27$	$3{,}883 \pm 0.07$	$8{,}059 \pm 27.41$
Scale Free (Cho et al., 2009)	1,000	$2{,}000 \pm 0.00$	$4{,}000 \pm 0.00$	$8{,}078 \pm 0.00$
Watts Strogatz	1,000	$2{,}000 \pm 0.00$	$4{,}000 \pm 0.00$	$8{,}078 \pm 0.00$

3.1.4. Dual Process Model

We designed a latent dual process model to simulate opinion formation. The first process determines, whether an agent even perceives the contribution. This is determined by affective value of the message (a_{message}). If it surpasses the noticing threshold (t_{noticing}), the content is processed. In the *personality model*, the *openness* has an influence on this threshold—more open agents will have lower thresholds.

The second process simulates opinion formation based on the latent process model. Each message has two components an affective ($m_{\text{affective}}$) and a cognitive value ($m_{\text{affective}}$). The geometric mean of those values with the agents existing internal affective ($a_{\text{affective}}$) and cognitive attitude ($a_{\text{cognitive}}$) is then compared against a behavioral threshold. If the process evokes a stronger "reaction" than the threshold the user adapts the attitude and will now forward the message once to all neighbors. In the *personality model*, the *conscientiousness* of the agent determines this threshold—more conscientious agents will have higher thresholds.

Data: t_{noticing} = the noticing threshold, $\bullet\, a_{\text{message}}$ = the affective value of the message
Result: determine whether to notice a message;
if $\underline{t_{\text{noticing}} < a_{\text{message}}}$ **then**
| evaluate for sending;
else
| ignore;
end

Algorithm 1: The noticing algorithm determines how a message gets noticed.

Data: the process values of agent and message
$\bullet\, t_{\text{sending}}$ = sending threshold $\bullet\, a_{\text{agent}}$ = affective value of agent $\bullet\, c_{\text{agent}}$ = cognitive value of agent
$\bullet\, a_{\text{message}}$ = affective value of message $\bullet\, c_{\text{message}}$ = cognitve value of message
Result: determine whether to send a message;
$a_{\text{totalvalue}} = \sqrt{a_{\text{agent}} \times a_{\text{message}}}$;
$c_{\text{totalvalue}} = \sqrt{c_{\text{agent}} \times c_{\text{message}}}$;
if $\underline{t_{\text{sending}} < \sqrt{a_{\text{totalvalue}} \times c_{\text{totalvalue}}}}$ **then**
| send;
else
| ignore;
end

Algorithm 2: The sending algorithm determines whether or not an agents forwards the message.

3.2. The Simulation Procedure

Using the aforementioned model components we ran 30,000 different simulations. We varied the network size (1,000, 2,000, 4039 agents), the network generator (see above), and the agent type (personality or random). The node with highest vertex centrality was chosen as the starting node, to spread messages, simulating the behavior of an opinion leader that introduces novel content to their "sub-network". We then ran the simulation until no more active senders were in the network. Agents that have sent the message, will not resend the message in later simulation steps. To reduce the impact of randomness we replicated each experiment 1,000 times using different random seeds. All experiments used a Mersenne Twister pseudo-random number generator. Initialization between different configurations of the experiments received the same random seed. Random seeds only varied for replications.

4. RESULTS

Using the agent-based model, we analyzed whether three different initial settings lead to different outcomes. As initial configurations, we first used different content types (*affective content*, *cognitive content*, *both*, *weak content*). Secondly, we used different network types (*Facebook*, *Barábasi-Albert*, *Watts Strogatz*, *Scale Free Network*, *Random Network*).

We see that the network generators behave relatively stable regarding network size (see **Figure 2**). Both clustering coefficient and community count are stable. There are differences between the networks though. Real data from facebook shows the largest cluster coefficient in all settings. Only the stochastic block model seems to capture a high clustering coefficient equally well. The Barábasi Albert model only leads to one large community, as in our case we used a preferential attachment generator, starting from one node.

Thirdly, we compared whether the use of a *personality model* for the creation of the agents in our simulation leads to a different outcome than the simulation runs without the *personality model*. Using these three different initial settings, we found some interesting results, that we show following.

The results of our simulation runs are depicted in **Figure 3**. The figure shows for each simulation step how many of the agents who saw the message also forwarded it. The number of forwarding agents is also visible for the six different network types (horizontal); for the agents with and without the use of the *personality model* (vertical); and the different content types (color). The first aspect we look at in the following is when the number of forwarding agents or online social media users in our simulation is highest or lowest.

4.1. Highest and Lowest Proportion of Forwarding Agents

As can be seen in this figure, never all agents have seen and forwarded the message. This applies to all initial settings. The number of forwarding agents was highest in the simulation (which is shown above right), where the content is *both*, affective and cognitive, where the agents have an according to the Big five factors designed *personality*, and where the agents are located in the *Watts Strogatz* network. Using these initial settings, more than 75% of the agents did forward the seen content.

In contrast, the lowest number of forwarding agents occurred in the simulation (which is shown above left), where the content is *weak* or *mostly cognitive*, where the agents are designed according to the *personality model* and where the agents are located in a *Barábasi Albert* network. The agents stop forwarding the message at the latest at the fourth simulation step and until then almost no agent has forwarded the message.

So far we considered, when the proportion of forwarding agents is highest or lowest. Following, we look at the single factors that could have an influence on the proportion of forwarding agents, starting with the four different content types, that are highlighted in different colors in the figure.

4.2. Content Types

Comparing the four different contents, most agents see and forward the content, that is *both* affective and cognitive. In every network type apart of the *Barábasi Albert* network, where we used the *personality model* (upper row), did in the end more than 75% forward the seem content. In the *Barábasi Albert* network still more than 50% forwarded the content. Without the *personality model* (lower row), still more than 60% forwarded the content in the four other network types and in the *Barábasi Albert* network did more than 30% forward the content. The agents forward the *mostly affective* content the second most and significantly more frequently than the other two contents.

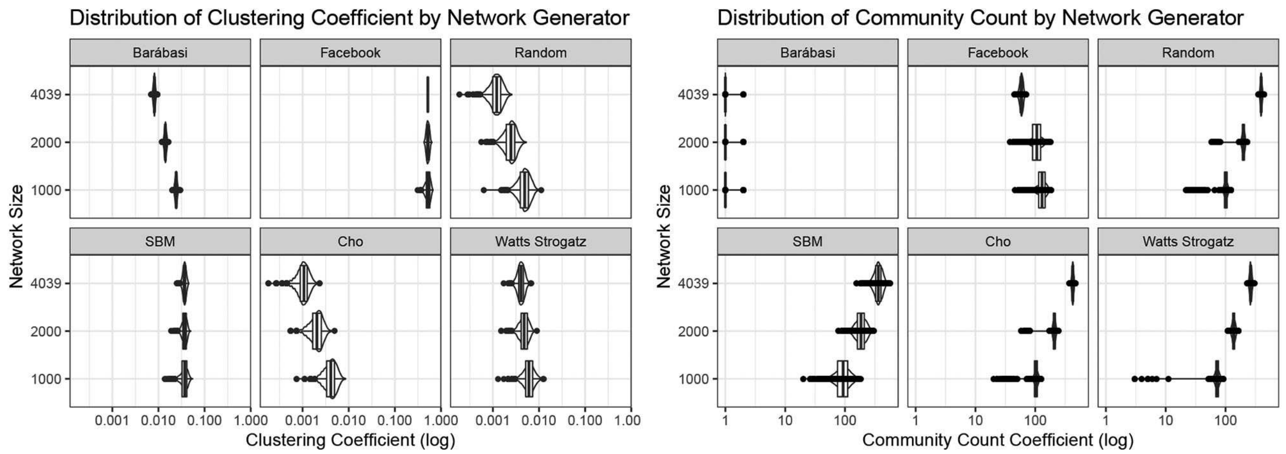


FIGURE 2 | Clustering coefficients and community counts for all network generators.

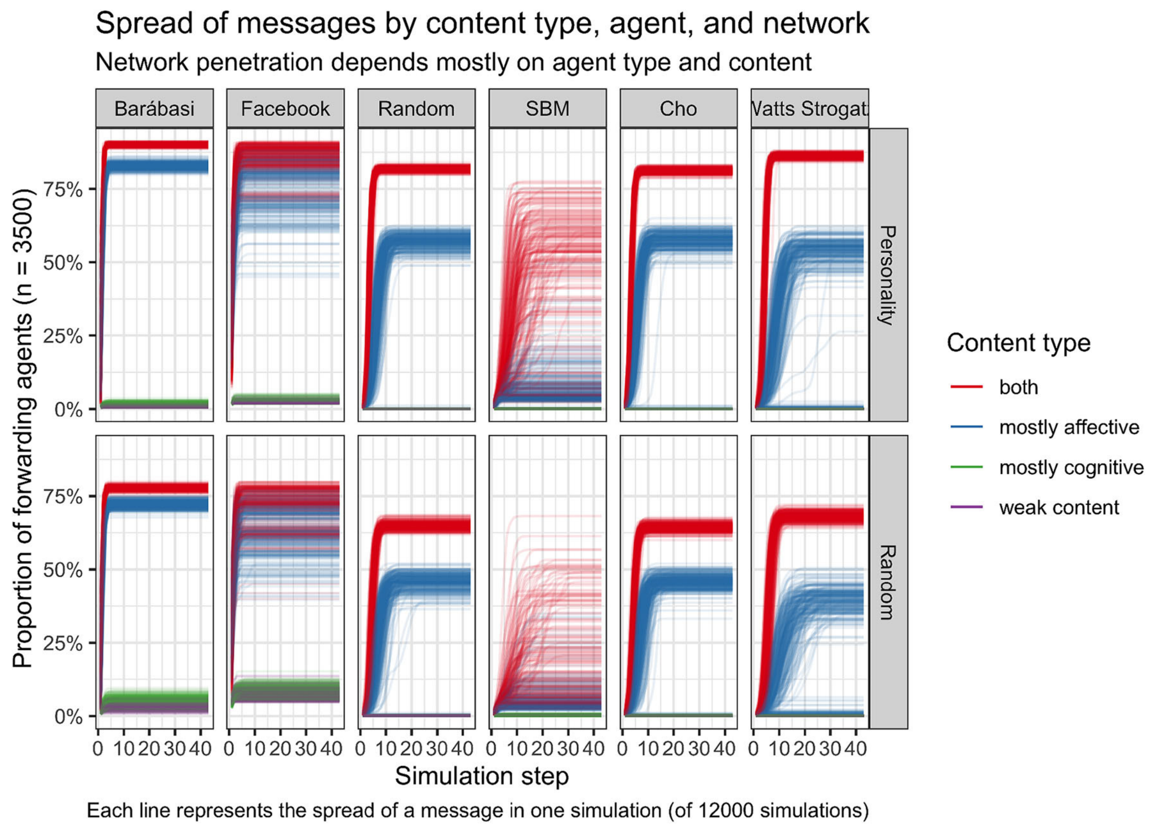


FIGURE 3 | Differences in content type become apparent in the spread of information.

The *weak content* as well as the *mostly cognitive content* are almost never forwarded at all. The agents also always stop forwarding the message before the eighth simulation step. Only with the *Facebook network* and without using the *personality model*, the *mostly cognitive content* is forwarded somewhat more frequently, but still forwarding does not exceed the eighth step of the simulation.

4.3. Network Type

After considering the influence of the content type, we now look at the different network types and how they influence the number of forwarding agents (horizontal). As can be seen in **Figure 3**, the proportion of forwarding agents differs only slightly between the *Random*, *Scale-Free*, and *Watts Strogatz* network. Hardly any difference can be seen between the red and blue

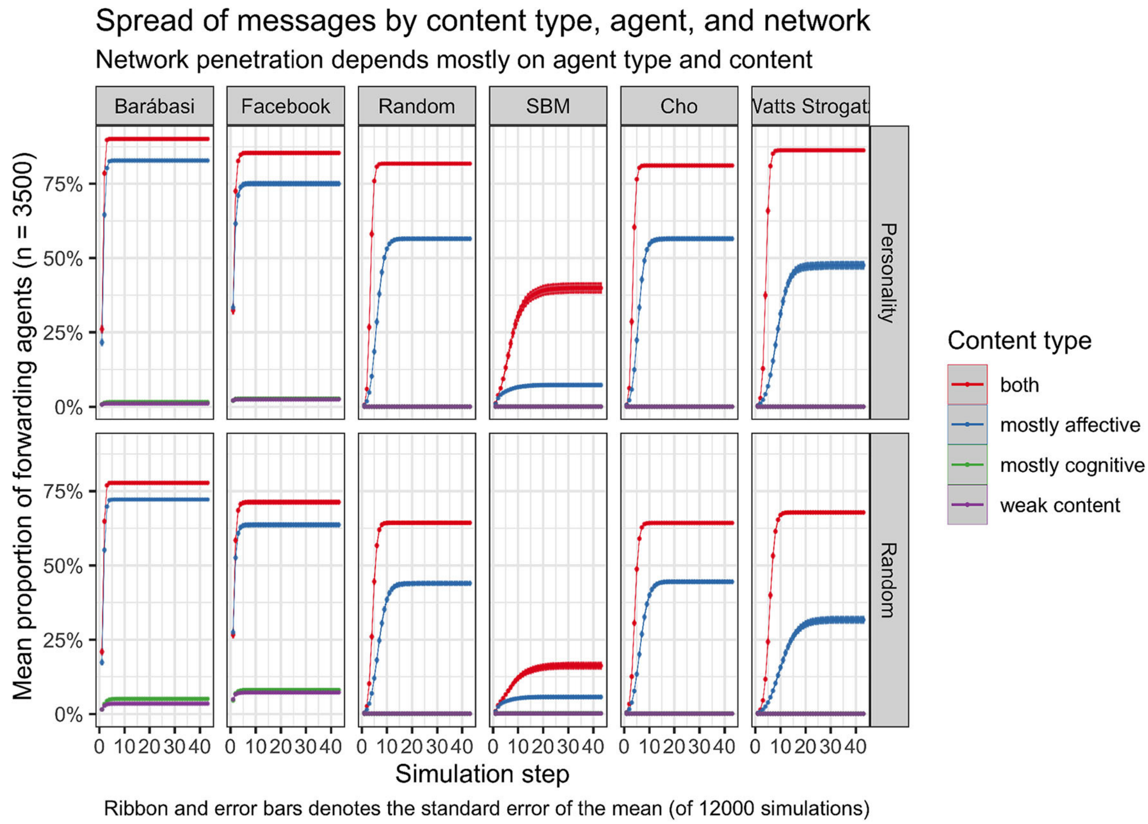


FIGURE 4 | The deviation from the mean spread of message is low across all simulation.

lines of the forwarding agents located either in the *Random* or in the *Scale-Free* network. The lines also look very similar for the agents in the *Watts Strogatz* network. While in the *Random* and in the *Scale-Free* network many agents already forward the *both content*, the number of forwarding agents for the *Watts Strogatz* network is a bit higher. If we look at the *mostly affective content*, the number of forwarding agents in the *Watts Strogatz* network differs more for the individual simulations than for the two previously mentioned networks.

Slightly larger differences can be seen for the *Barábasi Albert* and the *Facebook* network. In the *Facebook* and the *Barábasi Albert* network, the number of agents that forward the *both content* and the *mostly affective content* is more similar. In the *Barábasi Albert* network, compared to all other network types, fewer agents forward the two most forwarded contents.

4.4. Personality Model

Lastly, we compare the message spread in our simulations based on whether the personality of our agents followed a *personality model* or was *randomly* generated. **Figure 3** shows that the proportion of forwarding agents of the *mostly affective content* and the *affective and cognitive content* was always higher when they were equipped with a *personality model* in the *simulation*. The biggest deviation occurs when the agents are located in the *Barábasi Albert Network* or in the *Faceboook Network*. While in the *Barábasi Albert Network* around 20% of the agents forward the *affective and cognitive content*, when their personality is *randomly* generated, more than twice as many (around 44%) forward the *affective and cognitive content*, when their personality was designed through the *personality model*. Further, comparing the different ways of shaping the personality of the agents in the *Barábasi Albert Network*, the number of forwarding agents is almost identical for the other three contents.

In the *Faceboook Network*, the number of forwarding agents (with an intentionally created or random personality) is different for all contents. This type of network is also the only case where more agents with *random personality* forward the (*weak* and *mostly cognitive*) content than agents with an *intentionally created personality*. When the *personality model* was used, around 2% of the agents forwarded the *weak content* and around 3% forwarded the *mostly cognitive content*. Without the *personality model*, around 10% of the agents forwarded the *weak content* and around 11.5% of the agents forwarded the *mostly cognitive content*. In the end of the simulation runs, around 75% agents with a *randomly* generated personality forwarded the *affective and cognitive content*, with the *personality model*, the number of forwarding agents was around 87%. Around 62.5% of the agents with a *random* personality forwarded the *mostly affective* content and around 75% agents with an intentionally created personality forwarded the *mostly affective content*.

When the personality of the agents was *randomly* generated and the *Random*, *Scale-Free* and *Watts Strogatz* network was used, the number of forwarding agents (of the *cognitive and affective content*) remained below 75%. In contrast, the number was higher than 75% when we used the personality model. Likewise, the agents with an *intentionally created personality* forward the *mostly affective content* more often (around 44%), than the agents with *random personality* (around 25%). Regardless of how the personality of the agents is designed, no agent forwards the *mostly cognitive* and the *weak content*.

Overall, **Figure 3** shows that the proportion of forwarding agents mostly depends on the *content type* and if the agents have a *personality* designed according to the *personality model* or not. Even when looking at the standard error of the mean of proportions of agents that have seen or sent the message we see little deviation between the different graphs (see **Figure 4**) In contrast, the *network type* showed a lower influence except for *Barábasi Albert* and *Facebook* networks.

Lastly, **Figure 5** shows the results of a general linear model using agent type, network type and content type as predictors. As the predicted variable we used the number of agents that have sent the message until the last iteration step of each simulation. The model was significant with a null deviance of 473,939,679 on 11,999 degrees of freedom and a residual deviance of 93,791,997 on 11,990 degrees of freedom. The Akaike information criterion (AIC) for the final model was 141,644. This strengthens the importance of the message content in our model, but also highlights that the personality-based model contributed to modeling message spreading in our model. The only two factors not significant here were the network type being Cho or Watts Strogatz. As the null level of the network type the random graph was chosen. Our stochastic block model underestimated message spread compared against a real facebook model.

5. DISCUSSION

The first result of our study was that the generators that we have used did not behave exactly as real world network data. The focus in our study was not to perfectly simulate the network, but investigate the effect a dual process model in such networks. We found that all synthetic networks shows lower clustering coefficients, which may be derived from the processes and parameters we used. The ratio of nodes to edges was by far higher for real world data then for our synthetic networks. Tweaking of these parameters in the future will provide additional insights regarding the network generators.

Interestingly, the qualitative behavior of all networks was similar, yet with different ultimate levels of spread. The stochastic block model, chose for its ability to reproduce community like data in facebook, shows by far lesser spread then the real facebook data. The scale-free network using the algorithm proposed by Cho et al. (2009) behaves rather non-intuitive. The implementation was used without verification from the "LightGraphs" package. Future studies should check for faults in this implementation as well.

Moreover, we did not look fully into different network sizes. The slices we have used have no inherent value except for their readability. This is important to understand as the behavior of the graph limits is very well-understood. Our models are still far from the graph limits, and thus future work should verify the impact of network size in this model as well as phase-transitions in the network (critical states).

The results of our study show that in no simulation all agents saw the content in the end. In no case did more than 80% of the agents see the content. This means that the forwarding of the content inevitably stopped at some point, but what caused the agents to stop forwarding the content? Why was it sent so far nevertheless?

The results of our study indicate that the personality of an online social network user and the type of content have the greatest influence on the spread of messages, according to our simulations. In contrast, it makes almost no difference whether the agents interact in different network types.

We assumed, that varying the network type would effect the distribution of content depending on how individuals are connected. However, we were not able to observe this in our study. Therefore, it may be less crucial to explore how users are connected in social networks and, in contrast, more important to consider how users perceive different types of content and whether and how this depends on their personality. Nevertheless, as we have shown in section 2, the personality also influences how many people a user is connected to in social networks. In addition, a user can only see a content if he is connected in some way to a person who is sending it, which means that the importance of different network types should not be neglected. This finding comes with a caveat. Network generators have additional parameters that can be changed to create different node to edge ratios. In our configuration we always had far fewer edges than the real-life representation of facebook. In this particular network, it is interesting to see that also mostly cognitive appealing content was spread further in the network. Boundaries between content types are not as clear in this case.

In a network that is based on the Barábasi Albert network topology, less users see and forward a message. A possible explanation for this can be seen in the structure of this network. Because of the preferential attachment of Barábasi Albert, few users exist who are connected to a lot of other users and if one of these well-connected users chooses to reject a message, this has a greater effect on the whole forwarding process than a rejection of a single user in the other network types. By providing better interconnectedness between all nodes, this effect is diminished in other network types as there is still a good chance that the other users will receive the message from another user with whom they are also connected.

Contradicting to Nekovee et al. (2007) who found that small-world networks facilitate a very high initial message spread compared to random networks, there seems to be no difference between random and small word networks in message spread (Nekovee et al., 2007). We also expected that the Watts-Strogatz performs better because of the "Strength of Weak Ties theory" of Granovetter (1973). But other studies also found advantages of network structures with high local clustering relying on the complexity of the adoption and forwarding process (Centola, 2010).

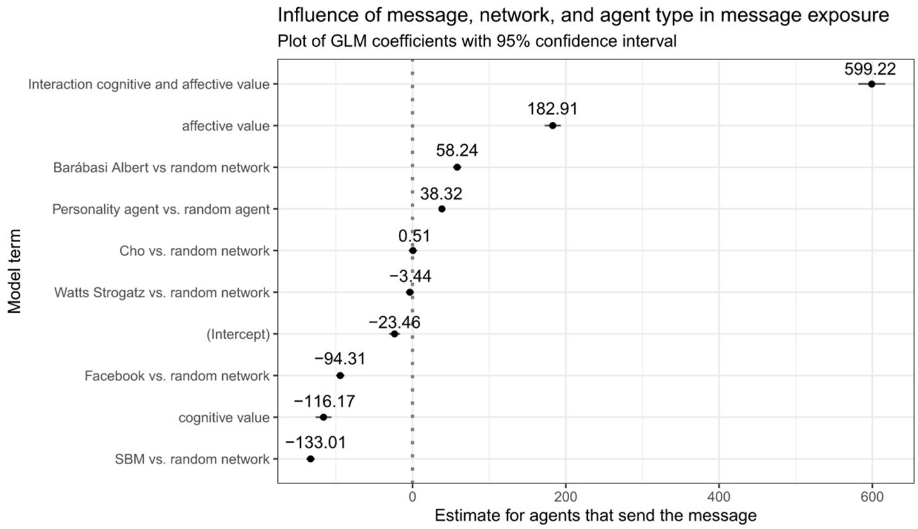


FIGURE 5 | Plotting the coefficients of a general linear model to determine the influence of network and agent type.

Confirmatory to the study of Xiong et al. the message spread of the scale-free networks and the Watts-Strogatz network as representative for small-world topologies is similar to each other (Xiong et al., 2015).

It has been shown that affective content is more likely to arouse readers' interest and also leads individuals to react to the content and forward it, for example (Oskamp and Schultz, 2005). Our results have also shown that the affective content is seen and shared much more frequently by social media users. Nevertheless, the combination of affective and cognitive content seems to increase the willingness to share a contribution the most.

As part of this study, we also designed the personality of the agents. In order to make the personality of the agents as realistic as possible and to resemble the personality of people or users of online social networks, we based our research on the results of the Big Five personality traits and how they relate to each other (see section 2.5.1). Thus, we designed the personality of the agents on the basis of the traits openness, conscientiousness and extraversion. Most interestingly, we see that including a personality model increases the reach of the message in all cases. This is partly due to the correlation of extraversion and openness in our model. More central nodes are more open and thus interact with more content. However, they are also more conscientious, but not sufficiently so to contain the spread of messages in a network. Integrating the personality perspective highlights the reach of both cognitively and affectively appealing messages. These show a very similar spread in many of the simulations.

By focusing this study on designing agent personality using three relevant features of the Big Five personality model, we omitted other features of the Big Five personality model and features of other personality models. This allowed us to design the relationships between the personality traits in a simple way and to design the personality of the agents in the model, which is always a simplification of the real world, in a sufficiently realistic way. Nevertheless, the personality of the agents can be designed more comprehensively. In the future, we would like to use the agreeableness and neuroticism of the Big Five personality model as well as other personality models to describe the personality of agents.

Further, we did not consider malicious individuals or *social bots* (Ferrara et al., 2016) in this simulation, although they have an influence on the spread of information in online social networks. Some studies (Bessi and Ferrara, 2016; Ferrara et al., 2016; Shao et al., 2018) showed, that social bots influence the public opinion by posting content and interacting with other social media users. Thus they behave like real social media users and are difficult to detect (Subrahmanian et al., 2016). Bots specifically send misinformation to users who are most likely to believe the information sent. This works well because people generally like to believe information that is popular or originates from their social environment (Jun et al., 2017). Using malicious agents/users or social bots in our simulation would have resulted in more spreaded weak content, what would have been interesting, but in this study (,as mentioned above,) we just concentrated on the influence of the three Big Five personality traits to keep the personality model relatively simple. In further studies we will extend the personality model not only by further personality traits, but also by different types of agents, such as malicious individuals, that try to manipulate the other agents in the simulation.

6. CONCLUSION AND OUTLOOK

With using an agent based model we found that the content type, the personality of a social media user and the type of network in which an individual is located have an influence on whether users see a contribution in the social network and whether they forward it. Overall, the willingness to forward a content depends more strongly on the content type and the personality of the agents than on the network type. This effect could even increase by using malicious agents or social bots. Still, there are special network types that have a great influence on how many users are reached by a content. The network type should therefore not be neglected in further research.

Since we saw that a network in which few users are connected to a lot of other users lead to a lower number of seeing and forwarding users, we want to consider network types with individual users connected to many other agents in the future. Here it could be particularly interesting to consider an agent of these well-connected agents as malicious.

Regarding the content type, a message, that combines affective and cognitive content increases the willingness to share the message the most. In contrast, social media users really do not want to share weak and mostly cognitive content no matter how the network is structured in our case. In the future it would also be interesting to take a closer look at the different forms of content. It is conceivable, for example, to compare different affective content to find out whether all affective content has a high probability to get forwarded or only content that appeals to certain emotions. It would also be interesting to look at different cognitively appealing contents or harmful content—e.g., click-bait, fake news, etc.—to find out what would make individuals forward that type of content more often than in our simulation or whether they actually never forward that content.

The integration of a personality model increases the willingness to forward content. At this point, however, we are not sure which other personality traits have the same effect and which personality traits could lead to the opposite effect and thus reduce the willingness to forward a contribution. It is conceivable, for example, that more conscientious persons would be even less likely to pass on harmful content. It is also conceivable that self-confident and extroverted social media as well as malicious users are more likely to forward content than people who fear negative feedback. In future, we will design the personality of our agents or social media users more comprehensively by including further personality traits other types of agents.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was funded by the State of North Rhine-Westphalia, Germany under the grant number 005-1709-0006, project Digitale Mündigkei and project number 1706dgn017.

REFERENCES

Adomavicius, G., and Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* 17, 734–749. doi: 10.1109/TKDE.2005.99

Alan, A. K., and Kabadayı, E. T. (2016). The effect of personal factors on social media usage of young consumers. *Proc. Soc. Behav. Sci.* 235, 595–602. doi: 10.1016/j.sbspro.2016.11.086

Albert, R., and Barabási, A. L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* 74, 47–97. doi: 10.1103/RevModPhys.74.47

Albi, G., Pareschi, L., and Zanella, M. (2016). Opinion dynamics over complex networks: kinetic modeling and numerical methods. *arXiv preprint arXiv:1604.00421*. doi: 10.3934/krm.2017001

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., and Stillwell, D. (2012). "Personality and patterns of facebook usage," in *Proceedings of the 3rd Annual ACM Web Science Conference, WebSci'12* (Evanston, IL). doi: 10.1145/2380718.2380722

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012). "The role of social networks in information diffusion," in *Proceedings of the 21st International Conference on World Wide Web, WWW '12* (New York, NY: ACM), 519–528. doi: 10.1145/2187836.2187907

Barrett, C. L., Beckman, R. J., Khan, M., Kumar, V. S. A., Marathe, M. V., Stretz, P. E., et al. (2009). "Generation and analysis of large synthetic social contact networks," in *Proceedings of the 2009 Winter Simulation Conference (WSC)* (New York, NY: IEEE), 1003–1014. doi: 10.1109/WSC.2009.5429425

Bessi, A., and Ferrara, E. (2016). Social bots distort the 2016 U.S. presidential election online discussion. *First Monday* 21. doi: 10.5210/fm.v21i11.7090

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: a fresh approach to numerical computing. *SIAM Rev.* 59, 65–98. doi: 10.1137/141000671

boyd, d. m., and Ellison, N. B. (2007). Social network sites: definition, history, and scholarship. *J. Comp. Mediat. Commun.* 13, 210–230. doi: 10.1111/j.1083-6101.2007.00393.x

Bromberger, S., Fairbanks, J., and other contributors (2017). *JuliaGraphs/LightGraphs.jl: LightGraphs*. doi: 10.5281/zenodo.889971

Burke, R. (2002). Hybrid recommender systems: survey and experiments. *User Model. User Adapt. Interact.* 12, 331–370. doi: 10.1023/A:1021240730564

Calero Valdez, A., and Ziefle, M. (2018). "Human factors in the age of algorithms. understanding the human-in-the-loop using agent-based modeling," in *Social Computing and Social Media. Technologies and Analytics*, ed G. Meiselwitz (Cham: Springer International Publishing), 357–371. doi: 10.1007/978-3-319-91485-5_27

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science* 329, 1194–1197. doi: 10.1126/science.1185231

Cheevasuntorn, K., Watanapa, B., Funilkul, S., and Krathu, W. (2017). "Factors influencing teenagers behavior on posting and sharing messages via Facebook," in *2nd International Conference on Information Technology (INCIT)* (Salaya Nakhonphathom), 1–4. doi: 10.1109/INCIT.2017.8257852

Cheng, J., Adamic, L., Dow, P. A., Kleinberg, J. M., and Leskovec, J. (2014). "Can cascades be predicted?" in *Proceedings of the 23rd International Conference on World Wide Web, WWW '14* (New York, NY: ACM), 925–936. doi: 10.1145/2566486.2567997

Cho, Y. S., Kim, J. S., Park, J., Kahng, B., and Kim, D. (2009). Percolation transitions in scale-free networks under the achlioptas process. *Phys. Rev. Lett.* 103:135702. doi: 10.1103/PhysRevLett.103.135702

Christakis, N. (2007). The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* 357, 370–379. doi: 10.1056/NEJMsa066082

Conte, R., Gilbert, N., Bonelli, G., Cioffi, C., Deffuant, G., Kertész, J., et al. (2012). Manifesto of computational social science. *Eur. Phys. J. Spcl. Top.* 214:325. doi: 10.1140/epjst/e2012-01697-8

Costa, P., and Mccrae, R. (1992). Neo PI-R professional manual. *Psychol. Assess. Resour.* 396.

Cullen, R., and Morse, S. (2011). "Who's contributing: do personality traits influence the level and type of participation in online communities," in *2011 44th Hawaii International Conference on System Sciences* (Kauai, HI), 1–11. doi: 10.1109/HICSS.2011.496

DeVito, M. A. (2017). From editors to algorithms. *Digital J.* 5, 753–773. doi: 10.1080/21670811.2016.1178592

Dong, Y., Yang, Y., Tang, J., Yang, Y., and Chawla, N. V. (2014). "Inferring user demographics and social strategies in mobile social networks," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14* (New York, NY: ACM), 15–24. doi: 10.1145/2623330.2623703

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., and Madden, M. (2014). *Social Media Update 2014*. Washington DC: Pew research center.

Eagly, A. H., and Chaiken, S. (1993). *The Psychology of Attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Epstein, J. M. (2007). *Generative Social Science: Studies in Agent-Based Computational Modeling (Princeton Studies in Complexity)*. Princeton, NJ: Princeton University Press.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17–60.

Facebook (2012). *Form S-1 Registration Statement*.

Ferrara, E., Varol, O., Davis, C. A., Menczer, F., and Flammini, A. (2016). The rise of social bots. *arXiv preprint arXiv abs/1407.5225*. doi: 10.1145/2818717

Fishbein, M., and Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Frees, B., and Koch, W. (2018). ARD/ZDF-onlinestudie 2018: zuwachs bei medialer internetnutzung und kommunikation. *Media Perspektiven* 9, 398–413. Available online at: https://www.ard-werbung.de/media-perspektiven/fachzeitschrift/2018/artikel/ardzdf-onlinestudie-2018-zuwachs-bei-medialer-internetnutzung-und-kommunikation/

Gilbert, N. (2008). *Agent-Based Models*. Los Angeles, CA: Sage. doi: 10.4135/9781412983259

Golbeck, J., Robles, C., and Turner, K. (2011). "Predicting personality with social media," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11* (New York, NY: ACM), 253–262. doi: 10.1145/1979742.1979614

Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *J. Pers. Soc. Psychol.* 59, 1216–1229. doi: 10.1037/0022-3514.59.6.1216

Gosling, S., Augustine, A., Vazire, S., Holtzman, N., and Gaddis, S. (2011). Manifestations of personality in online social networks: self-reported facebook-related behaviors and observable profile information. *Cyberpsychol. Behav. Soc. Network*. 14, 483–8. doi: 10.1089/cyber.2010.0087

Goyal, A., Bonchi, F., and Lakshmanan, L. V. S. (2008). "Discovering leaders from community actions," in *Proceedings of ACM 17th Conference on Information and Knowledge Management (CIKM)* (Napa Valley, CA). doi: 10.1145/1458082.1458149

Granovetter, M. S. (1973). The strength of weak ties. *Am. J. Sociol.* 78, 1360–1380. doi: 10.1086/225469

Grant, S., and Langan-Fox, J. (2007). Personality and the occupational stressor-strain relationship: the role of the big five. *J. Occup. Health Psychol.* 12, 20–33. doi: 10.1037/1076-8998.12.1.20

Guille, A., Hacid, H., Favre, C., and Zighed, D. A. (2013). Information diffusion in online social networks: a survey. *SIGMOD Rec.* 42, 17–28. doi: 10.1145/2503792.2503797

Gumelar, G., Akbar, Z., and Erik, E. (2018/07). "Engagement and the spread of fake news: personality trait as moderator," in *International Conference of Communication Science Research (ICCSR 2018)* (Surabaya: Atlantis Press). doi: 10.2991/iccsr-18.2018.34

Hartung, J. (2006). *Sozialpsychologie*. Kohlhammer.

Höllig, S., and Hasebrink, U. (2016). Nachrichtennutzung über soziale Medien im internationalen Vergleich. *Media Perspektiven* 11, 534–548.

Hughes, A., and Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *Int. J. Emerg. Manage.* 6, 248–260. doi: 10.1504/IJEM.2009.031564

Java, A., Kolari, P., Finn, T., Joshi, A., and Oates, T. (2007). Feeds That Matter: A Study of Bloglines Subscriptions. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2007). Computer Science and Electrical Engineering, University of Maryland, Baltimore County.

Jun, Y., Meng, R., and Johar, G. V. (2017). Perceived social presence reduces fact-checking. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5976–5981. doi: 10.1073/pnas.1700175114

Kang, M., Koch, C., and Pachón, A. (2015). The phase transition in multitype binomial random graphs. *SIAM J. Discrete Math.* 29, 1042–1064. doi: 10.1137/140973256

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5802–5805. doi: 10.1073/pnas.1218772110

L. DeFleur, M., and R.Westie, F. (1963). Attitude as a scientific concept. *Soc. Forces* 42, 17–31. doi: 10.2307/2574941

Leskovec, J., and Krevl, A. (2014). *SNAP Datasets: Stanford Large Network Dataset Collection*. Available online at: http://snap.stanford.edu/data

Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. (2009). Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. *Internet Math.* 6, 29–123. doi: 10.1080/15427951.2009.10129177

Li, M., Wang, X., Gao, K., and Zhang, S. (2017). A survey on information diffusion in online social networks: models and methods. *Information* 8:118. doi: 10.3390/info8040118

Lönnqvist, J.-E., and Itkonen, J. V. (2014). It's all about extraversion: why facebook friend count doesn't count towards well-being. *J. Res. Pers.* 53, 64–67. doi: 10.1016/j.jrp.2014.08.009

Lottridge, D., and Bentley, F. R. (2018). Let's hate together: how people share news in messaging, social, and public networks," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (New York, NY: ACM), 60:1–60:13. doi: 10.1145/3173574.3173634

Matsa, K. E., and Mitchell, A. (2014). *8 Key Takeaways About Social Media and News*. Washington, DC: Pew Research Center.

McAuley, J., and Leskovec, J. (2012). "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Lake Tahoe, NV: Curran Associates, Inc.) Vol. 1, 539–547.

McCombs, M. E., and Shaw, D. L. (1972). The agenda-setting function of mass media∗. *Public Opin. Q.* 36, 176–187. doi: 10.1086/267990

McGuire, W. J. (1985). "Attitudes and attitude change," in *Handbook of Social Psychology*, Vol. 9, 3rd Edn., eds G. Lindzey and E. Aronson (New York, NY: Random House), 233–346.

Meinefeld, W. (1977). Einstellung und soziales Handeln. *rororo-Studium: Sozialwissenschaft (Reinbek bei Hamburg: Rowohlt-Taschenbuch Verl.)* 1–236. Available online at: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-15735

Milgram, S. (1967). The small world problem. *Psychol. Tdy* 2, 60–67. doi: 10.1037/e400002009-005

Mossel, E., Neeman, J., and Sly, A. (2012). Stochastic block models and reconstruction. *arXiv preprint arXiv:1202.1499*.

Nekovee, M., Moreno, Y., Bianconi, G., and Marsili, M. (2007). Theory of rumour spreading in complex social networks. *Phys. A Stat. Mech. Appl.* 374, 457–470. doi: 10.1016/j.physa.2006.07.017

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U.S.A.* 98, 404–409. doi: 10.1073/pnas.98.2.404

Oskamp, S., and Schultz, P. W. (2005). *Attitudes and Opinions, 3rd Edn*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. doi: 10.4324/9781410611963

Pham, M.-D., Boncz, P., and Erling, O. (2013). "S3G2: a scalable structure-correlated social graph generator," in *Selected Topics in Performance Evaluation and Benchmarking*, eds R. Nambiar and M. Poess (Berlin; Heidelberg: Springer), 156–172. doi: 10.1007/978-3-642-36727-4_11

Power, R., and Pluess, M. (2015). Heritability estimates of the big five personality traits based on common genetic variants. *Transl. Psychiatry* 5:e604. doi: 10.1038/tp.2015.96

Rouchier, J., Cioffi, C., Polhill, J. G., and Takadama, K. (2008). Progress in model-to-model analysis. *J. Atif. Soc. Soc. Simul.* 11:8. Available online at: http://jasss.soc.surrey.ac.uk/11/2/8.html

Serrano, E., and Iglesias, C. A. (2016). Validating viral marketing strategies in twitter via agent-based social simulation. *Expert Syst. Appl.* 50, 140–150. doi: 10.1016/j.eswa.2015.12.021

Shang, Y. (2016). Groupies in multitype random graphs. *SpringerPlus* 5:989. doi: 10.1186/s40064-016-2705-4

Shang, Y. (2019). "Resilient consensus for expressed and private opinions," in *IEEE Transactions on Cybernetics*, 1–14. doi: 10.1109/TCYB.2019.2939929

Shao, C., Ciampaglia, G., Varol, O., Flammini, A., Menczer, F., and Yang, K.-C. (2018). The spread of low-credibility content by social bots. *Nat. Commun.* 9, 1–23. doi: 10.1038/s41467-018-06930-7

Stark, B., Magin, M., and Jürgens, P. (2017). *Ganz meine Meinung*?: *Informationsintermedi¨are* und Meinungsbildung - *eine Mehrmethodenstudie am Beispiel von Facebook, 55th Edn*. Düsseldorf: Landesanstalt für Medien Nordrhein-Westfalen (LfM).

Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., et al. (2016). The Darpa twitter bot challenge. *Computer* 49, 38–46. doi: 10.1109/MC.2016.183

Toscani, G., Tosin, A., and Zanella, M. (2018). Opinion modeling on social media and marketing aspects. *Phys. Rev. E* 98:022315. doi: 10.1103/PhysRevE.98.022315

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications, Vol. 8*. Cambridge University Press. doi: 10.1017/CBO9780511815478

Wasserman, S., and Pattison, P. (1996). Logit models and logistic regression for social networks: I. an introduction to Markov graphs and p*. *Psychometrika* 61, 401–425. doi: 10.1007/BF02294547

Watson, D., and Humrichouse, J. (2006). Personality development in emerging adulthood: Integrating evidence from self-ratings and spouse ratings. *J. Pers. Soc. Psychol*. 91, 959–974. doi: 10.1037/0022-3514.91.5.959

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440–442. doi: 10.1038/30918

Xiong, F., and Liu, Y. (2014). Opinion formation on social media: an empirical approach. *Chaos* 24:013130. doi: 10.1063/1.4866011

Xiong, F., Liu, Y., and Zhang, H. F. (2015). Multi-source information diffusion in online social networks. *J. Stat. Mech. Theory Exp*. 2015:07008. doi: 10.1088/1742-5468/2015/07/P07008

Zhang, Y., and Wu, Y. (2011). How behaviors spread in dynamic social networks. *Comput. Math. Organ. Theory* 18. doi: 10.1007/s10588-011-9105-7

Zheleva, E., Sharara, H., and Getoor, L. (2009). "Co-evolution of social and affiliation networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09* (New York, NY: ACM), 1007–1016. doi: 10.1145/1557019.1557128

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/frai.2020.00067

E-Commerce Shopping Motivation and the Influence of Persuasive Strategies

Ifeoma Adaji†, Kiemute Oyibo† and Julita Vassileva†*

Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:
Panagiotis Germanakos, SAP SE, Germany

Reviewed by:
Rosa Maria Vicari, Federal University of Rio Grande do Sul, Brazil
Dana Al Qudah, The University of Jordan, Jordan

***Correspondence:**
Ifeoma Adaji
ifeoma.adaji@usask.ca

†These authors have contributed equally to this work

Specialty section:
This article was submitted to AI for Human Learning and Behavior Change, a section of the journal Frontiers in Artificial Intelligence

Received: 21 October 2019
Accepted: 22 July 2020
Published: 23 November 2020

Citation:
Adaji I, Oyibo K and Vassileva J (2020) E-Commerce Shopping Motivation and the Influence of Persuasive Strategies. Front. Artif. Intell. 3:67. doi: 10.3389/frai.2020.00067

Persuasive strategies are used to influence the behavior or attitude of people without coercion and are commonly used in online systems such as e-commerce systems. However, in order to make persuasive strategies more effective, research suggests that they should be tailored to groups of similar individuals. Research in the traits that are effective in tailoring or personalizing persuasive strategies is an ongoing research area. In the present study, we propose the use of shoppers' online shopping motivation in tailoring six commonly used influence strategies: *scarcity, authority, consensus, liking, reciprocity,* and *commitment*. We aim to identify how these influence strategies can be tailored or personalized to e-commerce shoppers based on the online consumers' motivation when shopping. To achieve this, a research model was developed using Partial Least Squares-Structural Equation Modeling (PLS-SEM) and tested by conducting a study of 226 online shoppers. The result of our structural model suggests that persuasive strategies can influence e-commerce shoppers in various ways depending on the shopping motivation of the shopper. *Balanced buyers*—the shoppers who typically plan their shopping ahead and are influenced by the desire to search for information online—have the strongest influence on *commitment* strategy and have insignificant effects on the other strategies. *Convenience shoppers*—those motivated to shop online because of convenience—have the strongest influence on *scarcity*, while *store-oriented shoppers*—those who are motivated by the need for social interaction and immediate possession of goods—have the strongest influence on *consensus*. *Variety seekers*—consumers who are motivated to shop online because of the opportunity to search through a variety of products and brands, on the other hand, have the strongest influence on *authority*.

Keywords: persuasion, shopping motivation, e-commerce, shopper typology, persuasive strategies

INTRODUCTION

Simply selling products online is no longer sufficient for e-businesses to differentiate themselves from their online competitors. With many more companies now having an online presence, companies are seeking new ways to outdo their competitors. Businesses have to come up with new strategies to influence the purchasing decision of their clients.

Persuasion and how it is used to influence people's attitudes and the way they behave are an active research area in several domains including e-commerce. Persuasion is the use of influence strategies to change how people act and behave without coercion (Fogg, 2002). These strategies are often referred to as persuasive strategies (Fogg, 2002) and are implemented in various forms such as messages targeted at an audience. For example, some e-commerce companies use phrases such as "*Only a few left in stock*" to show that some products are in limited quantity. Existing research indicates that the use of persuasive strategies are more likely to result in a desired attitude or behavior change when these strategies are tailored to an individual or a group of individuals who are similar (Kaptein, 2011; Kaptein et al., 2012, 2015; Orji et al., 2014b).

Current efforts at tailoring persuasive strategies have used factors such as users' personality traits (Hirsh et al., 2012) and demographic data of users such as age (Phillips and Stanton, 2004), gender (Orji, 2016), and culture (Kramer and Spolter-Weisfeld, 2007). Despite the success in the use of personality, age, gender, and culture in tailoring persuasive strategies, in cases where these consumer characteristics are not known, such as in e-commerce, using these traits to tailor persuasive strategies is not possible. Therefore, for influence strategies to be personalized in online commerce, it is important to determine what other traits can be used to tailor persuasive strategies to individual users or groups of similar users to make them effective in bringing about a behavior or attitude change. We aim to fill this gap in the current paper by identifying if other factors such as a consumers' shopping motivation can be effectively used to tailor influence strategies to the consumers.

Research in e-commerce suggests that the intention of shoppers to buy a product is can be predicted by their motivation for shopping (Pappas et al., 2017). While shopping, online shoppers are not influenced the same way and thus, do not act the same way in terms of their shopping patterns and behaviors (Ganesh et al., 2010). Thus, in order to create a tailored or personalized online shopping experience for a shopper, it is essential to identify the factors that influence them (Pappas et al., 2017). Several typologies of shopping motivation exist. One such typology is that of Rohm and Swaminathan (2004), which classifies consumers into four categories according to their motivation for shopping online: *convenience shopper*, *store-oriented shopper*, *balanced buyer*, and *variety seeker*. We chose to use this typology in this study because of its popularity and widespread usage in e-commerce research (Ganesh et al., 2010; Pappas et al., 2017). Being able to identify what persuasive strategy each shopper type is influenced by could result in a shopping experience that is more personalized to the consumer. For instance, if *variety seekers* are influenced by *consensus* (looking to others who are similar to themselves in uncertainties) using messages that show consensus, for example, what products similar people have bought in the past, could influence this set of shoppers to buy particular products.

The aim of this paper is to identify what persuasive strategies e-consumers are influenced by based on their shopping motivation. To accomplish this, we conducted a study of 226 e-commerce shoppers to explore how the various shopper types (which are based on shopping motivation) are influenced by persuasive strategies. We measured persuasive strategies using Cialdini's six influence strategies (Cialdini, 2009) because they are commonly used in several domains including e-commerce (Kaptein and Parvinen, 2015). We developed a path model using partial least squares structural equation modeling (PLS-SEM) and tested it using the data from the survey. The result of our analysis suggests significant differences in the susceptibility of the various shopper types to the different influence strategies. In particular, *balanced buyers* were most highly influenced by *commitment* and were insignificantly affected by the other strategies. This suggests that *balanced buyers* are more likely susceptible to *commitment* strategy; thus, if they commit to purchasing a product, they will likely do so. Also, *convenience shoppers* were more influenced by *scarcity* compared to the other strategies, while *store-oriented shoppers* were more influenced by *consensus* compared to other strategies. Furthermore, *variety seekers* were more influenced by *authority* compared to other strategies. Possible guidelines in implementing these persuasive strategies in e-commerce are suggested.

RELATED WORK

Shopping Motivation

Research has shown that products can be effectively tailored to the various segments of consumers by classifying the customers according to how they are motivated to shop online (Rohm and Swaminathan, 2004). In addition, classifying consumers based on their motivation informs businesses of what clients look out for and their attitude during the shopping decision-making process (Keng Kau et al., 2003).

There are various taxonomies of online shoppers such as the typology of Keng Kau et al. (2003). They categorize e-commerce shoppers into six groups based on the information-seeking patterns of consumers in addition to their online motivation and concerns during the shopping process. Another popular typology is that of Rohm and Swaminathan (2004), who categorize online shoppers into four groups: *variety seekers, convenience shoppers*, *store-oriented shoppers*, and *balanced buyers* according to the shopping motivation of the consumers. According to the authors, the online convenience of shopping and the ability to save time and effort motivate *convenience shoppers* to shop online. This category of e-consumers, however, is not motivated to immediately acquire the products they buy. The possibility of searching for different brands and products from several stores motivates the *variety seekers.* Being able to explore product details online as the *variety seekers* motivates the *balanced buyers*. However, the *balanced buyers* differ from the *variety seekers* because the *balanced buyers* typically plan their purchases ahead, unlike the *variety seekers*, who do not. Social interaction motivates the *store-oriented shoppers*, in addition to the desire to acquire the purchased goods immediately.

The online clickstream data of consumers can be used to identify the various categories of shoppers. *Variety seekers*, for instance, compare different stores, products, and brands while shopping because they seek variety (Rohm and Swaminathan, 2004). *Variety seekers* will likely spend more time reviewing and comparing prices, promotions, brands, and the features of products before making a purchase decision (Keng Kau et al., 2003). Thus, if consumers' online click activity is analyzed, their browsing pattern can show if they are searching for a variety of products and if they can be classified as *variety seekers*. The *store-oriented shoppers* seek social interaction (Rohm and Swaminathan, 2004) and thus will likely engage in interaction or dialogue with other consumers on the e-commerce platform before making a purchase. Interaction in e-commerce is usually by asking other customers questions about the products they have previously purchased (Adaji and Vassileva, 2017) or by interacting with a site's chatbot if one exists. Thus, shoppers who typically interact with other consumers or with the site's chat agent before making purchases could be identified as *store-oriented shoppers*. In addition, because *store-oriented shoppers* are influenced to possess their products immediately (Rohm and Swaminathan, 2004), this category of shoppers will likely pay for express delivery of their products while other categories of shoppers will not. The online convenience of shopping and the ability to save time and effort motivate *convenience shoppers* to shop online (Rohm and Swaminathan, 2004). This category of consumers shops online for specific products and services; they do not seek variety across several channels but are motivated by the convenience of online shopping, effort, and time saving (Rohm and Swaminathan, 2004). Therefore, it is likely that *convenience shoppers* will not spend time and effort browsing different brands as the *variety seekers* would likely do. Their clickstream data could reveal their browsing patterns. Also, because social interaction does not influence *convenience shoppers*, this category of consumers may not participate on an e-commerce website's social platform, where questions are asked and answered and reviews posted. Furthermore, since *convenience shoppers* are not influenced to acquire purchased products immediately, they may be unwilling to pay extra for the express delivery of their products.

In the current paper, the typology of Rohm and Swaminathan (2004) was used because the four classes of shoppers are based on online shopping behavior and they have several similarities to other existing typologies, such as (Keng Kau et al., 2003; Moe, 2003). In addition, as far as we know, no other study exists that uses this popular typology in tailoring influence strategies in e-commerce.

Persuasive Strategies

According to Simons and Jones (2011) persuasion is "human communication designed to influence the autonomous judgments and actions of others." Persuasion attempts to change the way people think or act without being forced or coerced. Usually, with persuasion, the person being persuaded is in charge of the final decision of whether to change their behavior (Simons and Jones, 2011). Persuasive strategies are the different methods with which persuasion is implemented. Several taxonomies of persuasive strategies exist. The Persuasive Systems Design framework (PSD) (Oinas-Kukkonen and Harjumaa, 2008) consists of 24 persuasive strategies that the authors recommend for the design and development of persuasive systems. These are classified into four categories, defined by the task the strategy is intended to accomplish: primary task support, dialogue support, social support, and system credibility support. The categories of the PSD framework and the persuasive strategies that fall within each category are shown in **Table 1**.

TABLE 1 | Categories and persuasive strategies of the PSD framework.

Primary task support	Social support	System credibility support	Dialogue support
Reduction	Social learning	Trustworthiness	Praise
Tunneling	Social comparison	Expertise	Rewards
Tailoring	Normative influence	Surface credibility	Reminders
Personalization	Social facilitation	Real-world feel	Suggestion
Self-Monitoring	Cooperation	Authority	Similarity
Simulation	Competition	Third-party endorsement	Liking
Rehearsal	Recognition	Verifiability	Social role

The persuasive strategies of the PSD framework are commonly used in e-commerce systems to influence the shopping behavior of consumers. For example, amazon.com implements the 1-Click feature, which makes it easier for consumers to purchase items without having to go through the longer process of adding the item to their cart, filling out their shipping and payment details, and then placing the order for the product (Adaji and Vassileva, 2016). This significantly *reduces* the time it takes for a shopper to make a purchase. In addition, amazon.com allows its consumers to *self-monitor* their activities by providing a way for them to check the status of their orders and any previous purchases that they have made (Adaji and Vassileva, 2016). The online store childrensplace.com *suggests* other items to shoppers using the phrase "We think you'll also like" and images of suggested products. Walmart.ca influences people to shop by allowing them *learn* from others through the use of the "Questions and Answers" platform on the site.

Another common taxonomy of persuasive strategies is the six influence strategies of Cialdini which include *reciprocity, scarcity, commitment, authority, consensus, and liking* (Cialdini, 2009). *Reciprocity* is based on most people's need to always return a favor or repay in kind. An example of reciprocity is when an online bookstore offers its customers free e-books which could lead to more purchases from these customers because they feel the need to "return the favor[1]" A second example of *reciprocity* is the use of loyalty rewards programs offered by different companies. In their study of understanding customer retention and value based on their membership of a loyalty program, Bolton and Kannan (2000) conducted a study on a rewards-for-usage program offered by a financial services company. The company allows its members to accumulate points when they make purchases with their bank cards, which are redeemable through different stores offering a variety of products and services. The authors posit that the customers who benefited from the loyalty reward program were more likely to overlook the negative evaluations of the company because these customers believe they are receiving good value for their money in the form of the rewards program.

[1] *Exploiting the Power of Reciprocity*: Available online at: https://medium.com/@Omri_Yacubovich/exploiting-the-power-of-reciprocity-e214f96147c

Because humans are typically consistent in nature, when they commit to carry out a particular action, they usually do so. The *commitment* persuasive strategy suggests that if a system can get people to commit to a particular behavior, because of the consistent nature of humans, they likely carry out the target behavior (Cialdini, 2009). This strategy hinges on the theory of Cognitive Consistency, which suggests that because inconsistencies that are internal result in a state of tension in people, when faced with such internal inconsistencies, people behave in ways that could lower them (Feldman, 2013). Therefore, humans are commonly consistent in nature. In order to influence shoppers to commit to shopping with them, e-stores such as amazon.com offer consumers the opportunity to add products to a *wish list* (Kaptein, 2011). The clothing store childrensplace.com uses the foot in the door technique (Freedman and Fraser, 1966) by offering shoppers a discount on their next purchase.

The *consensus* persuasive strategy (also known as *social proof*) (Cialdini, 2009) proposes that people often took up to other people that they are similar to when not sure about how to behave and act. A common method of implementing *consensus* in e-commerce is by using the feature "customers who bought this item also bought," which displays products similar to that being viewed by a client. This feature is used on various e-commerce sites such as amazon.com, walmart.ca and realcanadiansuperstore.ca. Some online stores implement consensus by showing shoppers the number of people who have purchased a product (Kaptein, 2011).

According to Cialdini (2009), humans tend to believe and obey authority figures; therefore, when people decide what behavior to adopt in a given situation, the presence of authority figures can influence people's decisions. Authority figures include experts in a field, one's boss, or religious leaders (Cialdini, 2001, 2009). The endorsement of influencers and reviews from experts in a field are some ways that e-commerce companies implement *authority* (Kaptein, 2011).

Cialdini (2009) suggests that most times, people are more influenced by something or someone that they like; this describes the *liking* persuasive strategy. Therefore, if someone that a person likes makes a request, they are more likely to fulfill the request compared to a request from someone that the person does not like. Online consumers usually shop with companies that they like based on the recommendations and personalization that they receive from such companies (Li et al., 2013).

The *scarcity* principle, according to Cialdini (2009), is "the rule of the few." The author posits that humans crave for items that are limited and not readily available because scarce items are often considered more valuable than items that are abundant. In implementing scarcity, Cialdini suggests that businesses should highlight the unique benefits of a product, its exclusivity, and what people may lose by not purchasing a product (Cialdini, 2001). E-commerce vendors implement this strategy by announcing special limited time offers to their clients (Kaptein, 2011). Amazon.com implements scarcity by stating when a product is limited in stock or edition, with phrases like "only three left in stock." Laura.ca, a Canadian clothing retailer, uses the phrase "Hurry, *n* item(s) left for delivery," (where *n* represents a low number) in pink background to indicate a product is limited in stock. Walmart.ca uses the phrase "Almost sold out" in a red font to indicate items that are limited in quantity.

The use of Cialdini's six persuasive strategies to influence behavior change is an active research area. In their research on the effect of heterogeneity in persuasion in online systems, Kaptein and Eckles (2012) investigated three of Cialdini's six influence strategies: consensus, authority, and scarcity. Using product evaluations, the authors explored how the three persuasive strategies influence people differently. The authors concluded that, compared with a tailored approach, a one-size-fits-all method was less effective in influencing people to adopt a given behavior. In other words, the authors showed significant differences in the average effects of the three persuasive strategies. For example, some participants that were positively influenced by *consensus* were negatively influenced by *authority*. In addition, the authors suggested that using the wrong influence strategy could result in negative effects in terms of behavior change compared with using no strategy at all. Furthermore, using the best persuasive strategy for a person or similar individuals could influence them to carry out the desired change in attitude or behavior compared to using the best average strategy.

We chose to use Cialdini's six persuasive strategies in this study because they are popularly used in consumer studies research. In addition, compared with the PSD framework where some strategies are very similar to others (for example, simulation and rehearsal), the six strategies of Cialdini are very distinct and different from each other. Furthermore, there is currently no existing study that maps shoppers' online motivation to the persuasive strategies they are influenced by using Cialdini's strategies.

Tailoring Persuasive Strategies

Previous studies have shown that tailored persuasive strategies are more likely to bring about the desired behavior change compared to non-tailored strategies. For example, in their study of adaptive persuasive messages in e-commerce, Kaptein (2011) concluded that significant individual differences exist in users' responses to the implementation of various persuasive strategies. Similarly, in their study of influencing different gamer types, Orji et al. (2014b) determined that different gamer types are influenced by different persuasive strategies; the gamer type *achiever* is significantly influenced by *cooperation,* while a *daredevil* is influenced by *simulation*. Furthermore, Kaptein et al. (2012) studied the use of persuasive strategies in the form of messages to curtail snacking, and concluded that their study participants who received tailored messages significantly reduced their snacking consumption compared with the participants who did not. These results suggest that a one-size-fits-all approach to the implementation of persuasion will not likely bring about the desired behavior or attitude change among the users of a system.

Several factors have been used to tailor persuasive strategies. The use of personality traits is one of such factors. Hirsh et al. (2012), in their study of personalized persuasion, tailored persuasive messages for a single product via advertisements to shoppers based on their Big Five personality traits. The authors concluded that when persuasive messages are tailored to personality traits, it increased the impact of the messages. In their study of tailoring persuasion, Smith et al. (2016) tailored persuasive reminders to participants based on their personalities. The authors found significant differences in participants' preferences to the persuasive messages based on the participants' personalities. Alkiş and Taşkaya Temizel (2015) similarly researched the effect of tailoring influence strategies based on people's personalities. Their study of university students using the Big Five personality traits concluded that there were major differences in the influence of personality traits on influence strategies and thus that personality is a good factor in tailoring persuasive strategies.

The demographic data of users, such as age, gender, and culture, have also been used in personalizing influence strategies. In their study of motivational text messages, de Vries et al. (2017) concluded that gender influences the perception of motivational messages, thus, it can be used to tailor messages to people when the gender is known. Busch et al. (2016) investigated the role of gender in the persuasiveness of influence strategies. The authors concluded that different genders were influenced differently. Similarly, Orji et al. (2014a) examined the role of gender in the persuasiveness of influence strategies. The authors also concluded that males and females are influenced differently. Kramer and Spolter-Weisfeld (2007) researched the effect of the use of culture to tailor persuasive messages. Their results suggest that the cultural orientation of consumers significantly influenced their reception of personalized messages. The authors concluded that consumers, based on their culture—individualistic or collectivistic—responded differently to persuasive strategies. For example, collectivists were receptive to non-tailored recommendations, compared with individualists, who were not. Similarly, in her study of how the different cultures are influenced by persuasive strategies, Orji (2016) suggests that participants were influenced differently based on their culture, collectivistic or individualistic. Orji concluded that while collectivists were influenced by *reciprocity, authority, consensus*, and *liking*, individualists were not. Furthermore, Phillips and Stanton (2004) investigated age-related differences in persuasion and concluded that there are significant distinctions in the influence of persuasive strategies according to age. According to the authors, while younger consumers will likely recall information presented in ads, they will less likely be persuaded by it. On the other hand, older consumers will less likely recall information on ads but will more likely be persuaded by it.

In systems where these factors are not known, such as in e-commerce, it becomes difficult to tailor persuasive strategies to users to make these strategies more effective in bringing about the desired behavior change. For example, most e-commerce companies do not ask the gender or age of their clients during checkout. In addition, e-businesses make it possible for one to shop as a visitor without having to register an account with the merchant. Furthermore, people often shop for others, thus making it impossible to determine the gender of a shopper based on the content of their shopping cart. This study aims to fill this gap by using shoppers' online motivation instead of demographic data of shoppers to tailor persuasive strategies. There is currently no study that has done this to the best of our knowledge.

Other Factors That Influence Shopping Motivation

This paper focuses on the influence of persuasive strategies on shopping motivation, in particular, how different shoppers are influenced. We, however, recognize that other factors influence consumers' shopping motivation, such as the shopping value derived from the shopping experience. Value proposition has two popular dimensions: utilitarian and hedonic values. Consumers who possess high hedonic shopping value typically buy products for the happiness or pleasure that they get while shopping and not for how useful the product or service is (Overby and Lee, 2006; Bridges and Renée, 2008). Shoppers in this category are usually spontaneous, motivated to avoid pain, and drawn to pleasure (Babin et al., 1994; O'Shaughnessy and Jackson O'Shaughnessy, 2002).

Hedonic and utilitarian shopping values are an active research area in e-commerce. In their study of e-commerce consumers' purchase and shopping well-being, Yu et al. (2018) investigated the role of hedonic and utilitarian shopping values on the intention of consumers to purchase in shopping carnivals held online. The authors concluded that people with hedonic shopping values are persuaded by entertainment while those with utilitarian shopping values are influenced by saving money, selection, and convenience. Yu et al.'s study differs from that presented in the current paper because while the authors investigated shopping motivation in the form of hedonic and utilitarian shopping values while we investigated shopping motivation in the form of different shopper types.

Adaji et al. (2019) also researched the effect of influence strategies on the shopping motivation of online consumers based on their shopping value. The authors defined shopping motivation based on the value (hedonic or utilitarian) that shoppers derived while shopping. The authors suggest that people with high hedonic value are persuaded to purchase scarce and limited products while people with high utilitarian shopping value are influenced by their social circles. The present study differs from that of Adaji et al. because the authors defined shopping motivation based on the hedonic and utilitarian shopping values of consumers but the present study defines shopping motivation based on the shopper type taxonomy of Rohm and Swaminathan (2004). To the best of our knowledge, this has not been done before.

RESEARCH DESIGN AND METHODOLOGY

The research question, design and methods used in addressing the research question are presented in this section.

Research Question

The overarching research question that is addressed by this paper is the following:

How are e-commerce shoppers influenced by persuasive strategies based on their different motivations to shop online?

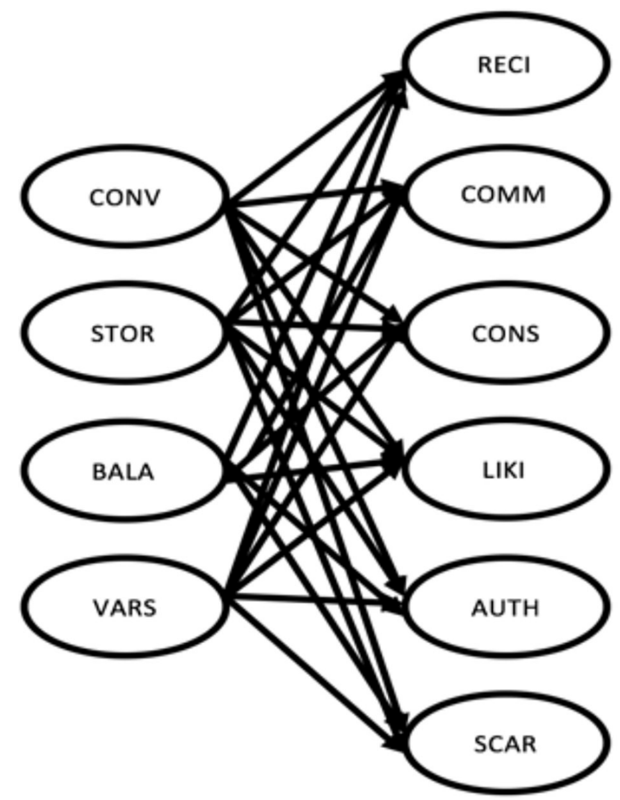


FIGURE 1 | Research model with all paths assumed to be positive. STOR, store-oriented shopper; CONV, convenience shopper; VARS, variety seeker; BALA, balanced buyer; RECI, reciprocity; COMM, commitment; CONS, consensus; LIKI, liking; AUTH, authority; SCAR, scarcity.

Methodology: Structural Measurement Model

To answer our research question, we developed a path model (shown in **Figure 1**) using PLS-SEM to measure the susceptibility of the four shopper types (based on online shopping motivation: *variety seekers, convenience shoppers, store-oriented shoppers, and balanced buyers*) to Cialdini's (2009) six influence strategies: *scarcity, consensus, authority, commitment, reciprocation, and liking*. The model was developed using four constructs to represent the four shopper types and six constructs to represent the six persuasive strategies. As defined by the research question, the aim of the model is to measure the influence of the different persuasive strategies on the shopper types—in other words, to determine which persuasive strategy has the highest influence on the different shopper types.

Rohm's scale which consists of four constructs and 17 questions was used to measure shopping motivation (Rohm and Swaminathan, 2004). The susceptibility to persuasive strategies was measured using the *susceptibility to persuasive strategies* scale of Kaptein et al. (2009), which is made up of six constructs and 32 questions.

In carrying out the PLS-SEM, bootstrapping was implemented using a random sample size of 5,000 (with replacement) to derive the distribution to be used in the model for the different constructs as suggested by Hair et al. (2016). Also, we determined the indicator reliability of our model, its internal consistency reliability, the convergent validity and discriminant validity to ensure they met the minimum requirements as required in PLS-SEM analysis (Hair et al., 2016). These results are presented in section Evaluation of Global Measurements. The path coefficient, β, between constructs was also computed.

To test the model, we created a survey online using the instruments mentioned above. We measured all items on a seven-point Likert scale, where 1 was strongly disagree, and 7 was strongly agree.

TABLE 2 | Participants' demographics.

Demographics	Value	Frequency (%)
Age	Below 30	55
	Between 30 and 49 inclusive	40
	Above 50	5
Gender	Female	44
	Male	56
Size of household	1–3 people inclusive	63
	4–5 people inclusive	34
	6 or more people	4
Combined Income of household	Below US$30,000	40
	Between US$30,000 and $75,000	42
	More than US$75,000	18
Origin/Continent	Europe	8
	Asia	35
	North America	48
	Others	9

Participants

We carried out a study of e-commerce shoppers to test our model. The questions were presented in an online survey. In all, 226 e-commerce shoppers were recruited to take part in the study. Recruitment was done using Amazon's Mechanical Turk (AMT). In addition, we recruited some participants through various online social media and the news board of our University. We used AMT because it allows one to recruit a diverse set of participants, and it is an accepted method of recruiting participants (Hirsh et al., 2012; Jia et al., 2016). We have successfully used online social media and news boards in the past with success (Busch et al., 2016). Therefore, we used them again to recruit participants in this study. Participants were asked to answer the questions in the context of grocery shopping. The Behavioral Ethics Board of our University approved the study. The demographics of our participants are presented in **Table 2**.

DATA ANALYSIS AND RESULTS

We analyzed the survey data using the SmartPLS tool[2]. SmartPLS is a commonly used tool for PLS-SEM and is popularly used in

[2]https://www.smartpls.com/

the research community because of its ease of use and ease of interpretation of results (Wong, 2013; Hair et al., 2016).

Partial Least Squares Structural Equation Modeling (PLS-SEM)

PLS-SEM is used mainly in exploratory research to develop theories. It focuses on describing the variance of dependent variables in a research model. Even with a small sample size, PLS-SEM is known to achieve significant statistical results and does not require the distributional assumptions of other statistical methods (Hair et al., 2016). PLS-SEM does not rely on any distributional assumptions. Rather it uses bootstrapping to derive a distribution to be used in the model. In bootstrapping, subsamples are selected randomly and replaced from the original dataset. This goes on repeatedly until a substantial number of random samples have been created (Hair et al., 2016).

In carrying out the analysis of the structural model, bootstrapping was implemented with a random sample size of 5,000 (with replacement) as recommended by Hair et al. (2016).

Evaluation of Global Measurements

Research (Hair et al., 2016) suggests that the relationship between indicators (which are measures of a construct or the questions asked for each construct) of each construct should be evaluated before the relationship between the constructs is considered. This is achieved by computing a model's internal consistency reliability, indicator reliability, convergent validity, and discriminant validity (Hair et al., 2016). The results of these measurements are presented in the following section.

Internal Consistency Reliability

The use of Cronbach's alpha in assessing internal consistency reliability is not recommended since it assumes that all the indicators of a construct are equally reliable (Hair et al., 2016). This does not always happen because oftentimes, the indicators of a construct do not have the same outer loadings. In addition, the number of items on a scale influences Cronbach's alpha; an increase in Cronbach's alpha often results from an increase in the number of items (Hair et al., 2016). A commonly used alternative for measuring internal consistency that, researchers suggest, is better than Cronbach's alpha is composite reliability (Wong, 2013; Hair et al., 2016). Composite reliability indicates whether the indicator variables (the questions asked for each construct) are a good measure of a construct. **Table 3** shows that the composite reliability of all constructs is >0.6, the acceptable threshold (Hair et al., 2016). Therefore, we conclude that among all constructs, high levels of composite reliability were established.

Convergent Validity

The degree of correlation between the indicators of a construct is referred to as the convergent validity. Because the indicators of a construct are alternatives to measuring the same construct, they should share a high variance. In structural equation modeling, the convergent validity of a model is often measured with the average variance extracted (AVE) (Wong, 2013; Hair et al., 2016). **Table 3** shows that the constructs in the model have the minimum acceptable AVE values of at least 0.5 (Wong, 2013; Hair et al., 2016).

TABLE 3 | Composite reliability and AVE of constructs.

Constructs	Composite reliability	Average variance extracted (AVE)
Convenience shopper	0.875	0.637
Store oriented shopper	0.816	0.60
Balanced buyer	0.863	0.677
Variety seeker	0.638	0.50
Reciprocity	0.897	0.638
Scarcity	0.789	0.50
Authority	0.868	0.569
Commitment	0.832	0.50
Consensus	0.860	0.607
Liking	0.853	0.537

Indicator Reliability

Indicator reliability describes the size of the relationship between indicators that make up a construct and the construct (Hair et al., 2016). Research suggests that this relationship, known as the outer loadings, should be at least 0.4 for exploratory studies (Hulland, 1999; Wong, 2013; Hair et al., 2016). As shown in **Table 4**, the outer loadings in the model meet this criterion.

Discriminant Validity

Discriminant validity defines the extent to which a model's constructs differ from each other. Establishing discriminant validity indicates that each construct in the model is unique (Fornell and Larcker, 1981; Wong, 2013; Hair et al., 2016). If the square root of the AVE for each construct is higher than its highest correlation with other constructs, one can conclude that discriminant validity is established (Wong, 2013; Hair et al., 2016). As shown in **Table 5**, the square root of the AVE in bold is greater than the correlation values in each row. Therefore, we conclude that discriminant validity is established.

Structural Measurement Model: Evaluation

The structural model's results show the relationship between the independent variable and the dependent variable and how strong this relationship is. In addition, the results of the structural model describe how much the variances of the independent variables are defined by the dependent variables. This is represented by the path coefficients, β, between constructs. **Table 6** shows the results of our structural model. The number of asterisks which range from 1 to 4 indicates how significant each direct path is. The asterisks represent the $p < 0.05$, < 0.01, < 0.001, and < 0.0001, respectively.

Balanced buyer is the most strongly affected by the strategy *commitment* ($\beta = 0.327$), and other strategies have insignificant effects. This suggests that *balanced buyers* are likely susceptible to a *commitment* strategy. *Convenience shopper* is the most strongly affected by *scarcity*, while *consensus* has the strongest effect on

TABLE 4 | Outer loadings of model.

	Convenience shopper	Store-oriented shopper	Balanced buyer	Variety seeker	Reciprocity	Scarcity	Authority	Commitment	Consensus	Liking
Convenience shopper 1	0.855									
Convenience shopper 2	0.853									
Convenience shopper 3	0.715									
Convenience shopper 4	0.759									
Store-oriented shopper 1		0.651								
Store-oriented shopper 2		0.806								
Store-oriented shopper 3		0.848								
Store-oriented shopper 4		0.721								
Balanced buyer 1			0.824							
Balanced buyer 2			0.810							
Balanced buyer 3			0.834							
Balanced buyer 4			0.808							
Variety seeker 1				0.643						
Variety seeker 2				0.726						
Variety seeker 3				0.771						
Variety seeker 4				0.698						
Variety seeker 5				0.701						
Reciprocity 1					0.814					
Reciprocity 2					0.846					
Reciprocity 3					0.860					
Reciprocity 4					0.670					
Reciprocity 5					0.785					
Scarcity 1						0.638				
Scarcity 2						0.768				
Scarcity 3						0.695				
Scarcity 4						0.769				
Scarcity 5						0.717				
Authority 1							0.715			
Authority 2							0.772			
Authority 3							0.833			
Authority 4							0.728			
Authority 5							0.715			
Commitment 1								0.695		
Commitment 2								0.683		
Commitment 3								0.634		
Commitment 4								0.788		
Commitment 5								0.724		
Commitment 6								0.788		
Consensus 1									0.727	
Consensus 2									0.735	

(Continued)

TABLE 4 | Continued

	Convenience shopper	Store-oriented shopper	Balanced buyer	Variety seeker	Reciprocity	Scarcity	Authority	Commitment	Consensus	Liking
Consensus 3									0.798	
Consensus 4									0.806	
Consensus 5									0.775	
Consensus 6									0.703	
Liking 1										0.731
Liking 2										0.776
Liking 3										0.720
Liking 4										0.709

TABLE 5 | Correlation of constructs.

Constructs	Convenience shopper	Store-oriented shopper	Balanced buyer	Variety seeker	Reciprocity	Scarcity	Authority	Commitment	Consensus	Liking
Convenience shopper	**0.798**									
Store-oriented shopper	−0.273	**0.775**								
Balanced buyer	0.287	0.089	**0.822**							
Variety seeker	0.165	0.180	0.243	**0.707**						
Reciprocity	0.295	0.076	0.251	0.262	**0.799**					
Scarcity	0.251	0.155	0.091	0.246	0.232	**0.707**				
Authority	0.223	0.149	0.245	0.345	0.522	0.338	**0.754**			
Commitment	0.300	0.085	0.431	0.281	0.589	0.256	0.513	**0.707**		
Consensus	0.076	0.269	0.038	0.265	0.298	0.264	0.429	0.264	**0.779**	
Liking	0.207	0.172	0.083	0.301	0.366	0.307	0.535	0.402	0.583	**0.733**

The bold diagonal shows the square roots of AVE.

TABLE 6 | Path coefficients of the structural model.

Shopper types	Authority	Commitment	Consensus	Liking	Reciprocity	Scarcity
Balanced buyer	0.116 n.s.	**0.327******	−0.078 n.s.	−0.062n.s.	0.126 n.s.	−0.054 n.s.
Convenience shopper	0.186**	0.203**	0.138 n.s.	0.240**	0.259****	**0.295***
Store-oriented shopper	0.142*	0.084 n.s.	**0.276******	0.200*	0.105 n.s.	0.209*
Variety seeker	**0.260****	0.153 n.s.	0.211***	0.240**	0.170*	0.173 n.s.

N.s., not significant. The number of asterisks which range from 1 to 4 indicates how significant each direct path is. The asterisks represent the p-values < 0.05, < 0.01, < 0.001 and < 0.0001 respectively. The bold values show what persuasive strategy has the highest influence on the various shopper types.

store-oriented shopper. In addition, *authority* has the strongest effect on v*ariety seeker*.

DISCUSSION

This study aims to identify what persuasive strategy each shopper type is influenced by. To answer the research question "How are e-commerce shoppers influenced by persuasive strategies based on their different motivations to shop online?" our results indicate that there are significant differences in the effects of various persuasive strategies on e-commerce shoppers as a result of their online shopping motivation. For example, while *balanced buyers* are influenced by *commitment* ($\beta = 0.327$), *store-oriented shoppers* have the strongest susceptibility to *consensus* ($\beta = 0.276$).

Balanced Buyers

The ability to search online for information motivates the *balanced buyers*, who are similar to the *variety seekers* (Rohm and Swaminathan, 2004). On the contrary, *balanced buyers* do not typically schedule their purchases in advance and are likely to make impulse purchases online (Rohm and Swaminathan, 2004). The results of this study suggest that *balanced buyers* are only influenced by the *commitment* strategy ($\beta = 0.327$). The *commitment* strategy (Cialdini, 2009) suggests that people are naturally consistent. Thus, if people commit to carrying out a target behavior, because of their consistent nature, they will likely carry out the behavior. Therefore, if an e-commerce site can get *balanced buyers* to commit to a particular behavior or action, this could result in this category of shoppers carrying out that behavior because they are influenced by *commitment*. This suggests that if *balanced buyers* commit to shopping for healthful meals, for example, they will likely do so.

Cialdini (2001) suggests that a choice made explicitly, voluntarily, and publicly is more likely to change one's behavior compared to one made implicitly. An example of *commitment* is the "Foot in the door" technique (Freedman and Fraser, 1966). It suggests that if a person agrees to, and carries out a small request, it increases the likelihood that they will carry out a similar larger request. An example of implementing *commitment* in e-commerce is when an e-commerce company offers consumers a discount on their next purchase as shown in **Figure 2**.

This suggests that in tailoring persuasive strategies to shoppers in e-commerce, where the age, gender, and culture of shoppers are usually unknown, the shopping motivation of the consumer can be used.

Convenience Shoppers

The minimal effort involved in online shopping, in addition to convenience and the time it saves compared to traditional shopping, motivates the *convenience shoppers* (Rohm and Swaminathan, 2004). These consumers do not expect to receive their goods immediately and are not motivated to carry out any social interaction while shopping. Furthermore, they do not search for a variety of products from different retailers (Rohm and Swaminathan, 2004). Our results suggest that *scarcity* ($\beta = 0.295$) has the strongest influence on *convenience shoppers*.

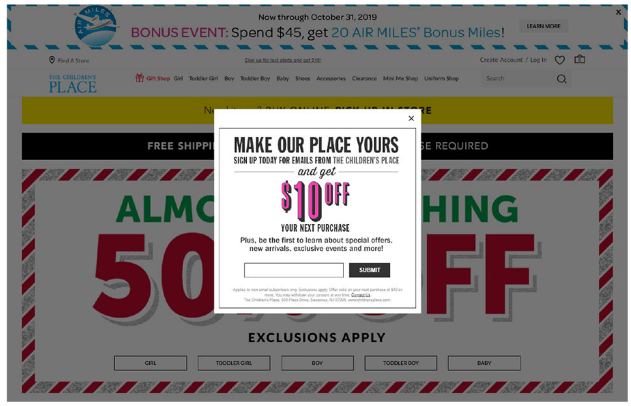


FIGURE 2 | Example of the commitment strategy.

Because this category of shoppers does not search for variety, it is not surprising that they are influenced by items that are limited.

In implementing *scarcity*, Cialdini (2001) suggests that one highlight the unique benefits of an item and, in addition, state its exclusivity. E-commerce companies implement *scarcity* by stating when a product is *limited in stock*, is a *rare item*, or a *limited-edition* item. For example, Amazon[3] uses the phrase "*n* items in stock" (where *n* represents a low number) when they are running out of an item. As shown in **Figure 3A**, Laura[4], a popular clothing retailer in Canada, uses the phrase "Hurry, *n* item(s) left for delivery" (where *n* represents a low number) in pink background (indicated by the yellow arrow) when a product is limited in stock. Walmart[5], a popular North American multinational corporation, uses the phrase "Almost sold out" in a red font as indicated by the yellow arrow in **Figure 3B**.

This result indicates that for shoppers in e-commerce, since the consumers' demographic data such as their age, gender, and culture are not known, their shopping motivation is a good factor in deciding how to tailor persuasive strategies.

Store-Oriented Shoppers

The desire to possess their products immediately and social interaction motivate *store-oriented* consumers to shop online (Rohm and Swaminathan, 2004). Our results suggest that this category of shoppers has the strongest influence on the persuasive strategy *consensus* ($\beta = 0.276$). *Consensus* (also referred to as social proof) implies that people look to others who are similar to them for suggestions on how to behave, especially when in doubt (Cialdini, 2001). This finding is reasonable because *store-oriented shoppers* are motivated to shop by social interaction. Thus, it is possible that they look to others for answers to questions about products and purchase decisions when they are shopping,

Cialdini (2001) suggests that in implementing the *consensus* strategy, one could use peer power whenever it is available. For example, he suggests that reviews from satisfied customers work

[3]https://www.amazon.ca/
[4]https://www.laura.ca/
[5]https://www.walmart.ca/

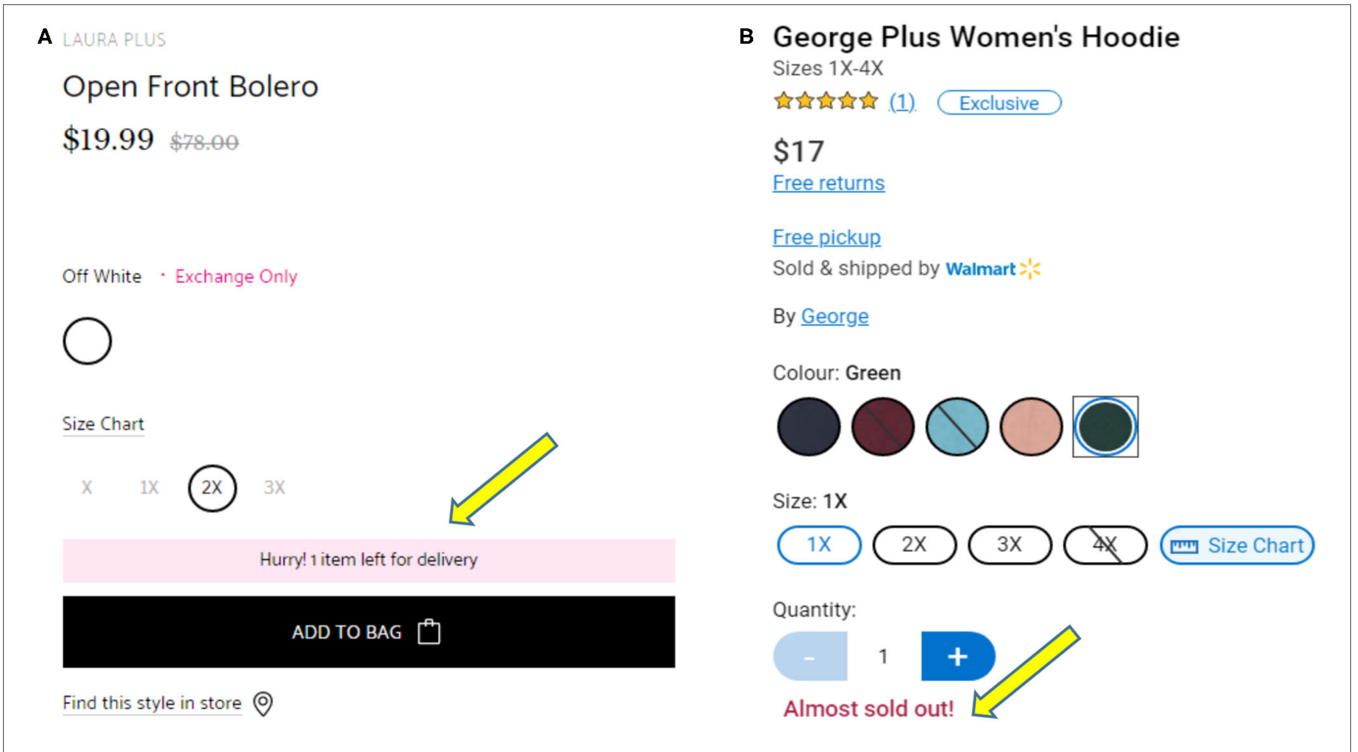


FIGURE 3 | (A) Example of implementing scarcity in www.laura.ca. **(B)** Example of implementing scarcity in www.walmart.ca.

better to influence prospective customers when the prospective client and satisfied client have something in common. One way to implement *consensus* in e-commerce is to show shoppers what products other consumers have bought or to show the products that are often purchased together. As shown in **Figure 4**, Amazon uses the phrase "Customers who read this also read" to show what books others have purchased based on the content of one's shopping cart.

In tailoring persuasive strategies to shoppers in e-commerce, this result shows that the shopping motivation of the consumer can be used.

Variety Seekers

Variety seekers are motivated by the desire to seek a variety of products across various stores, product types, and brands (Rohm and Swaminathan, 2004). Our results suggest that this category of shoppers is most strongly influenced by the persuasive strategy *authority* ($\beta = 0.260$). This result is plausible because *variety seekers* who compare products across various channels will likely come across several reviews from *experts* who are knowledgeable about the product.

The notion behind the *authority* strategy is that people listen to experts more than they listen to non-experts (Cialdini, 2001). Thus, claiming that a statement is one from experts could make people such as *variety seekers*, who are influenced by the *authority* strategy, change their attitude or behavior. Factors that can trigger the *authority* principle include (1) the use of titles such as Dr., Prof., CEO, (2) clothes such as religious outfits worn by priests, monks, and nuns, (3) status symbols such as an expensive car or suit (Cialdini, 2009), and (4) as well as quotes and endorsements from experts and authority figures. One way to implement *authority* while presenting a product to consumers is by using messages such as "The ministry of healthy suggests five daily servings of fruit" to influence consumers to purchase more fruit. Another example is to show reviews of people in authority such as book reviews of prominent authors or reviewers. As shown in **Figure 5**, Amazon includes book reviews from authority figures such as the Wall Street Journal.

Our results are an indication that the shopping motivation of consumers can be used as a factor in tailoring persuasive strategies to make them more effective in bringing about a change in attitude or behavior.

The Strategies to Implement for the Various Shopper Types

The result shown in **Table 6** indicates that *commitment* is the only strategy positively and significantly associated with *balanced buyers*. This suggests that consumers in this category will be significantly influenced only by *commitment*, making it the best strategy to implement for *balanced buyers*. *Convenience shoppers*, on the other hand, are influenced by all strategies except *consensus*, with *scarcity* having the strongest influence. *Store-oriented shoppers* are significantly influenced by *authority, consensus, liking*, and *scarcity*, with *consensus* having the strongest influence, while *variety seekers* are influenced by *authority, consensus, liking,* and *reciprocity,* with *authority* being the strongest.

FIGURE 4 | Example of implementing consensus on amazon.ca.

"The novel buzzes with the energy of numerous adventures, love affairs, [and] twists of fate." (*The Wall Street Journal*)

FIGURE 5 | Example of implementing authority on amazon.ca.

Best General Strategy for Shopper Types

For system designers who want to implement persuasive strategies based on the shopper types, if the designer's objective is an overall average effect across all shopper types, we recommend two strategies. The first recommended strategy is *liking*. Only two strategies, *liking* and *authority*, significantly influence three of the four shopper types. However, the influence on the shopper types of *liking* is stronger than the effect of *authority* for almost all the strategies. Therefore, *liking* is a better overall strategy to implement across all shopper types compared with *authority* or the other strategies. The second recommended strategy is *commitment*. No other strategy has an influence on *balanced buyers* except *commitment*. Thus, if a system designer is implementing strategies that will include all shopper types including *balanced buyers*, *commitment* has to be implemented in addition to *liking*.

If, on the other hand, the design objective is to maximize the effect of the persuasive strategy on the individual shopper types, the recommended strategies are *commitment, scarcity, consensus*, and *authority* for *balanced buyers, convenience shoppers, store-oriented shoppers*, and *variety seekers,* respectively.

Limitations

This study is limited in a few ways. First, the results are self-reported and do not depend on the direct observation of participants. This is, however, common practice in consumer-based research as many successful studies in the past have been self-reported. Second, the sample size, 226, represents only a fraction of e-commerce shoppers worldwide. However, we are, of the opinion that with the thorough analysis we have carried out and the results obtained in this paper, the results would likely be similar if we had more participants.

CONCLUSION AND FUTURE WORK

Research suggests that influence strategies are effective in bringing about a change in people's attitudes and behavior. However, to make them effective, persuasive strategies should be tailored to people with similarities. In e-commerce, where the gender and age of shoppers are not known to the e-commerce vendor, there is a need to identify other traits that are effective in tailoring persuasive strategies to make them more effective in changing shoppers' attitudes or behavior. To fill this gap, this paper aimed to investigate how influence strategies could be tailored to e-commerce shoppers according to how they (consumers) are motivated to shop online. In particular, the paper aimed to answer the research question How are e-commerce shoppers influenced by persuasive strategies based on their different motivations to shop online? To achieve this, a structural model was developed using

PLS-SEM and was evaluated by carrying out a study of 226 online shoppers.

Our results contribute and advance research in the area of e-commerce personalization and tailoring of persuasive strategies by showing that the different types of shoppers are significantly influenced by persuasive strategies differently. To answer our research question, different shopper types are influenced differently. Thus, a one-size-fits-all approach where the same persuasive strategies are applied to all types of shoppers will likely not be effective in changing shoppers' behavior. Rather, tailoring persuasive strategies to individual shopper types will result in the desired behavior change. In particular, the *commitment* strategy had the highest influence on the *balanced buyer* shopper type while the other strategies had insignificant effects. This indicates that *balanced buyers* are susceptible to *commitment* compared to other strategies. This implies that if balanced buyers can commit to making a purchase, they will likely carry it out. The "foot in the door" technique is one way that e-commerce companies can influence balanced buyers to commit to shopping with them by offering special discounts on their next purchase. *Convenience shoppers* had the highest influence on *scarcity*, which suggests that products that are labeled as limited, scarce, or rare will likely be more attractive to *convenience shoppers. Store-oriented shoppers* were most highly influenced by *consensus*, which suggests that, when in doubt, *convenience shoppers* look to others in their social circle for what to buy. This implies that by highlighting the products that others in their social circles have purchased, the shopping decision of *convenience shoppers* can be influenced. *Variety seekers*, on the other hand, were most highly affected by the influence strategy *authority*. This suggests that *variety seekers* can be influenced to purchase products because of people in authority.

These results suggest guidelines for the implementation of persuasive strategies by e-commerce platforms to make these persuasive strategies more effective in influencing the purchasing decisions of shoppers. For example, in a bid to make people shop for more healthful foods when shopping online, an e-commerce platform can present healthful foods that are limited in edition, rare, or scarce to *convenience shoppers* because this category of shoppers is influenced to purchase products that are limited, rare, or scarce.

Although we are limited by a small sample size, we chose to use PLS-SEM in our study because PLS-SEM performs well even with small samples. We are still in the process of data collection and will repeat the study with more participants in the future. In addition, we will implement and test these results on an online shopping site in the future. In the proposed study, the strategies identified will be implemented for the different shopper types and the reactions of shoppers to these strategies will be noted and compared to the results presented in this study.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Saskatchewan Human Ethics Review Board. The patients/participants provided their informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was partially supported by the NSERC Discovery grant of JV.

REFERENCES

Adaji, I., Oyibo, K., and Vassileva, J. (2019). "Effect of shopping value on the susceptibility of E-commerce shoppers to persuasive strategies and the role of gender," in *International Conference on Persuasive Technology* (Cyprus), 270–282. doi: 10.1007/978-3-030-17287-9_22

Adaji, I., and Vassileva, J. (2016). "Evaluating personalization and persuasion in e-commerce," in *International Workshop on Personalized Persuasive Technology* (Salzburg), 107–113.

Adaji, I., and Vassileva, J. (2017). "Perceived effectiveness, credibility and continuance intention in E-commerce. a study of amazon," in *Proceedings of 12th International Conference on Persuasive Technology* (Amsterdam), 293–306. doi: 10.1007/978-3-319-55134-0_23

Alkiş, N., and Taşkaya Temizel, T. (2015). The impact of individual differences on influence strategies. *Personal. Individ. Differ*. 87, 147–152. doi: 10.1016/j.paid.2015.07.037

Babin, B. J., Darden, W. R., and Griffin, M. (1994). Work and/or fun: measuring hedonic and utilitarian shopping value. *J. Consum. Res*. 20, 644–656. doi: 10.1086/209376

Bolton, R., and Kannan, P. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *J. Acad. Mark. Sci*. 28, 95–108. doi: 10.1177/0092070300281009

Bridges, E., and Renée, F. (2008). Hedonic and utilitarian shopping goals: the online experience. *J. Bus. Res*. 61, 309–314. doi: 10.1016/j.jbusres.2007.06.017

Busch, M., Mattheiss, E., Reisinger, M., Orji, R., Fröhlich, P., and Tscheligi, M. (2016). More than sex: the role of femininity and masculinity in the design of personalized persuasive games," in *Persuasive Technology. PERSUASIVE 2016. Lecture Notes in Computer Science, Vol. 9638,* eds A. Meschtscherjakov B. De Ruyter V. Fuchsberger M. Murer, and M. Tscheligi (Cham: Springer), 219–229. doi: 10.1007/978-3-319-31510-2_19

Cialdini, R. (2001). Harnessing the science of persuasion. *Harvard Bus. Rev*. 79, 72–81. doi: 10.1038/scientificamerican0201-76

Cialdini, R. B. (2009). *Influence: Science and Practice*. Pearson Education Boston.

de Vries, R. A. J., Truong, K. P., Zaga, C., Li, J., and Evers, V. (2017). A word of advice: how to tailor motivational text messages based on behavior change theory to personality and gender. *Personal Ubiquitous Comput*. 21, 675–687. doi: 10.1007/s00779-017-1025-1

Feldman, S. (2013). *Cognitive Consistency: Motivational Antecedents and Behavioral Consequents.* Academic Press.

Fogg, B. J. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do.* Ubiquity. doi: 10.1145/764008.763957

Fornell, C., and Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104

Freedman, J. L., and Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique 3. *J. Personal. Soc. Psychol.* 4, 155–202. doi: 10.1037/h0023552

Ganesh, J., Reynolds, K., Luckett, M., and Pomirleanu, N. (2010). Online shopper motivations, and e-store attributes: an examination of online patronage behavior and shopper typologies. *J. Retail.* 86, 106–115. doi: 10.1016/j.jretai.2010.01.003

Hair, J. Jr., Hult, T., Ringle, C., and Sarstedt, M. (2016). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Sage Publications. doi: 10.15358/9783800653614

Hirsh, J. B., Kang, S. K., and Bodenhausen, G. V. (2012). Personalized persuasion: tailoring persuasive appeals to recipients' personality traits. *Psychol. Sci.* 23, 578–581. doi: 10.1177/0956797611436349

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies. *Strate. Manag. J.* 20, 195–204. doi: 10.1002/(SICI)1097-0266(199902)20:2<195::AID-SMJ13>3.0.CO;2-7

Jia, Y., Xu, B., Karanam, Y., and Voida, S. (2016). "Personality-targeted gamification: a survey study on personality traits and motivational affordances," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, CA), 2001–2013. doi: 10.1145/2858036.2858515

Kaptein, M. (2011). "Adaptive persuasive messages in an e-commerce setting: the use of persuasion profiles," in *European Conference on Information Systems* (Helsinki), 183.

Kaptein, M., De Ruyter, B., Markopoulos, P., and Aarts, E. (2012). Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transact. Intell. Syst.* 2, 1–25. doi: 10.1145/2209310.2209313

Kaptein, M., and Eckles, D. (2012). Heterogeneity in the effects of online persuasion. *J. Interact. Mark.* 26, 176–188. doi: 10.1016/j.intmar.2012.02.002

Kaptein, M., Markopoulos, P., de Ruyter, B., and Aarts, E. (2009). "Can you be persuaded? individual differences in susceptibility to persuasion," in *IFIP Conference on Human-Computer Interaction* (Berlin; Heidelberg: Springer), 115–118.

Kaptein, M., Markopoulos, P., De Ruyter, B., and Aarts, E. (2015). Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. *Int. J. Human Comput. Stud.* 77, 38–51. doi: 10.1016/j.ijhcs.2015.01.004

Kaptein, M., and Parvinen, P. (2015). "Dynamically adapting sales influence tactics in e-commerce," in *Marketing Dynamism & Sustainability: Things Change, Things Stay the Same....* (Cham: Springer), 445–454.

Keng Kau, A., Tang, Y. E., and Ghose, S. (2003). Typology of online shoppers. *J. Consum. Mark.* 20, 139–156. doi: 10.1108/07363760310464604

Kramer, T., and Spolter-Weisfeld, S. (2007). The effect of cultural orientation on consumer responses to personalization. *Mark. Sci.* 26, 246–258. doi: 10.1287/mksc.1060.0223

Li, Y., Liang, T., and Wei, K. (2013). "How can personalized online services affect customer loyalty: the relationship building perspective," in *Service Science and Innovation.* (Kaohsiung). doi: 10.1109/ICSSI.2013.25

Moe, W. (2003). Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. *J. Consum. Psychol.* 13, 29–39. doi: 10.1207/153276603768344762

Oinas-Kukkonen, H., and Harjumaa, M. (2008). "A systematic framework for designing and evaluating persuasive systems," in *Proceedings of International Conference on Persuasive Technology* (Berlin; Heidelberg), 164–176. doi: 10.1007/978-3-540-68504-3_15

Orji, R. (2016). "The impact of cultural differences on the persuasiveness of influence strategies," in *Proceedings of the 11th International Conference 2016 on Persuasive Technology* (Salzburg), 38–41.

Orji, R., Mandryk, R., and Vassileva, J. (2014a). "Gender and persuasive technology: examining the persuasiveness of persuasive strategies by gender groups," in *International Conference on Persuasive Technology* (Padua), 48–52.

Orji, R., Vassileva, J., and Mandryk, R. L. (2014b). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Model. User Adapt. Interact.* 24, 453–498. doi: 10.1007/s11257-014-9149-8

O'Shaughnessy, J., and Jackson O'Shaughnessy, N. (2002). Marketing, the consumer society and hedonism. *Eur. J. Mark.* 36, 524–547. doi: 10.1108/03090560210422871

Overby, J. W., and Lee, E.-J. (2006). The effects of utilitarian and hedonic online shopping value on consumer preference and intentions. *J. Bus. Res.* 59, 1160–1166. doi: 10.1016/j.jbusres.2006.03.008

Pappas, I., Kouurouthanassis, P., Giannakos, M., and Lekakos, G. (2017). The interplay of online shopping motivations and experiential factors on personalized e-commerce: a complexity theory approach. *Telemat. Inform.* 34, 730–742. doi: 10.1016/j.tele.2016.08.021

Phillips, D. M., and Stanton, J. L. (2004). Age-related differences in advertising: Recall and persuasion. *J. Target. Measure. Analy. Mark.* 13, 7–20. doi: 10.1057/palgrave.jt.5740128

Rohm, A. J., and Swaminathan, V. (2004). A typology of online shoppers based on shopping motivations. *J. Bus. Res.* 57, 748–757. doi: 10.1016/S0148-2963(02)00351-X

Simons, H. H. W., and Jones, J. (2011). *Persuasion in Society*. Taylor & Francis.

Smith, K., Dennis, M., and Masthoff, J. (2016). "Personalizing reminders to personality for melanoma self-checking," in *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (Halifax, NS), 85–93. doi: 10.1145/2930238.2930254

Wong, K. (2013). Partial least squares structural equation modeling (PLS-SEM) techniques using SmartPLS. *Mark. Bull.* 24, 1–32.

Yu, H., Zhang, R., and Liu, B. (2018). Analysis on consumers' purchase and shopping well-being in online shopping carnivals with two motivational dimensions. *Sustainability* 10:4603. doi: 10.3390/su10124603

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership