Designing Digital Health Interventions for Sedentary Behavior Change

Doctoral thesis for obtaining the academic degree Doctor of Doctor of Natural Sciences (Dr.rer.nat.)

submitted by
Wang, Yunlong

at the

Universität Konstanz

Faculty of Sciences
Department of Computer and Information Science

Konstanz, 2019
Doctoral Thesis of the University of Konstanz
Date of the oral examination: 24\textsuperscript{th} October 2019

1. Reviewer: Prof. Dr. Harald Reiterer
2. Reviewer: Prof. Dr. Falk Schreiber
ACKNOWLEDGEMENTS

First of all, I want to thank my supervisor, Prof. Dr. Harald Reiterer, who gave me invaluable guidance and support throughout my Ph.D. research. He has led me to the HCI research field and showed me what a decent professor should be. Also, I want to thank other professors, colleagues, and co-authors, who helped me in the past five years. They are Prof. Dr. Falk Schreiber, Prof. Dr. Britta Renner, Dr. Bjoern Sommer, Dr. Laura Koenig, Dr. Roman Raedle, Dr. Jens Mueller, Dr. Simon Butscher, Daniel Klinkhammer, Johannes Zagermann, Maximillian Duerr, Corinna Breitinger, Dr. Lingdan Wu, Dr. Ahmed Fadhil, to name a few.

Second, I would like to give many thanks to my parents, my sister, and other family members for their endless support and love. Besides, I would like to thank my fiancé, Ke Wang, for her sweet accompany and love. I also appreciate the warm support from all my friends in Konstanz and other places in Germany, as well as those in China.

Last but not least, I want to acknowledge the financial support from the China Scholarship Council. I was lucky to hold the scholarship, which enabled me to focus on my research interest when pursuing my Ph.D. degree. Also, I was pleased to join the SMARTACT project and get financial support for conducting my studies.


ABSTRACT

Sedentary lifestyles become increasingly ubiquitous along with the development of industry and technology, due to the reduced necessity of physical activity during work. Unfortunately, evidence has shown that prolonged sedentary behavior is a risk factor to many chronic diseases (e.g., obese, type 2 diabetes, cardiovascular diseases, and colon cancers), independently from moderate- and vigorous-intensity physical activity. Therefore, it is urgent to improve the sedentary lifestyle to prevent these diseases.

Comparing with traditional intervention programs, digital health interventions have several advantages: (1) they are easier to scale up, (2) they could be more lightweight and cheaper, and (3) they could enable real-time and context-aware interventions. Therefore, in this dissertation, we focus on the technologies of digital health interventions for sedentary behavior change. We developed and evaluated new technologies based on daily-use digital devices (i.e., smartphones and PCs) for sedentary behavior change, as well as explored the potential impact of the emerging digital devices (i.e., augmented-reality head-mounted displays) on users’ movement behavior.

Due to the interdisciplinary nature of this research field, digital health (eHealth), this dissertation involves theories and practice from health psychology, physiology, public health, human-computer interaction, and data mining. In the beginning, we conducted a systematic review on technologies for sedentary behavior change at work. Through this review, we noticed several limitations in prior work, including the inconsistency of using taxonomies and the lack of applying behavioral theories. Therefore, we then proposed a holistic framework, TUDER (Targeting, Understanding, Designing, Evaluating, and Refining), for designing and reporting digital health interventions by integrating existing taxonomies and theories from different communities.

Following our proposed framework, we conducted two intervention studies for reducing the sedentary behavior of office workers and college students. In the first study, we developed a mobile app – SedVis – visualizing users’ mobility and sedentary patterns extracted by our proposed data mining algorithm to support action planning of sedentary behavior change. The results showed that using novel visualizations of mobility patterns might be a promising avenue for reducing sedentary time by facilitating self-monitoring of the behavior and providing engaging feedback.

Our second intervention study aimed to evaluate two context-aware PC reminders, the point-of-choice prompt and our proposed SedBar (an always-on progress bar), regarding the effects on users’ sedentary behavior change and perceived usefulness and interruption. This study suggested that context-aware PC reminders hold great potential for sedentary behavior change. The point-of-choice prompt might be a promising tool for reducing sedentary time for screen-
based workers. Users perceived the SedBar as useful, but the effectiveness should be further investigated.

Besides light-weight and practical software applications on smartphones and PCs, we also explored the potential impact of the augment-reality head-mounted displays (e.g., HoloLens) on office workers’ movement behavior. By comparing the participants’ movement behavior under different levels of movement freedom and flexibility when performing tasks using HoloLens, we provided several design implications for future applications on AR-HMDs from the health perspective.

Lastly, in the dissertation, we stepped out of the scope of sedentary behavior change and overviewed the research on general health behavior change in the HCI community. Following our proposed holistic framework of digital health interventions, we discussed the patterns and trends in the field of health behavior change research in HCI. Based on the reviewed studies, we provided implications and research opportunities for future work.

Im Vergleich zu herkömmlichen Interventionen bieten digitale Gesundheitsmaßnahmen mehrere Vorteile: (1) sie sind einfacher zu skalieren, (2) sie sind einfacher und günstiger und (3) sie ermöglichen kontextsensitive Interventionen in Echtzeit. Daher konzentrieren wir uns in dieser Dissertation auf die Technologien digitaler Gesundheitsinterventionen zur Veränderung des Bewegungsapparates. Wir haben neue Technologien entwickelt und evaluiert, die auf digitalen Geräten für den täglichen Gebrauch (z. B. Smartphones und PCs) zur Änderung des Sitzverhaltens beitragen, und die möglichen Auswirkungen der aufkommenden digitalen Geräte (z. B. kopfgesteuerte Augmented-Reality-Displays) auf unseren Bewegungsverhalten analysiert.


Gemäß unserem vorgeschlagenen Konzept führten wir zwei Interventionsstudien durch, um das sitzende Verhalten von Büroangestellten und Studenten zu verringern. In der ersten Studie haben wir mit SedVis eine mobile App entwickelt, die die Beweglichkeit und Bewegungsmuster der Benutzer visualisiert, die mit unserem vorgeschlagenen Data-Mining-Algorithmus ermittelt wurden, um die Aktionsplanung zur Änderung des Bewegungsmusters zu unterstützen. Die Ergebnisse zeigten, dass die Verwendung neuartiger Visualisierungen von Mobilitätsmustern ein vielversprechender Weg sein könnte, um die Ruhezeit zu verkürzen, indem die
Selbstüberwachung des Verhaltens erleichtert und ein ansprechendes Feedback gegeben wird.


Schließlich sind wir in der Dissertation aus dem Bereich der Sitzverhaltensänderung herausgetreten und haben einen Überblick über die Forschung zur allgemeinen Veränderung des Gesundheitsverhaltens in der HCI-Gemeinschaft gegeben. In Anlehnung an unseren vorgeschlagenen ganzheitliches Konzept für digitale Gesundheitsinterventionen diskutierten wir die Muster und Trends im Bereich der Forschung zu Gesundheitsverhaltensänderungen bei HCI. Basierend auf den überprüften Studien liefern wir Implikationen und Forschungsmöglichkeiten für zukünftige Forschung.
Parts of this research were previously published in the following publications. Reused material is indicated in the beginning of each chapter where applicable. The asterisks indicate the co-first authorship.

Journal Publications


Conference & Workshop Papers


ACM Workshop on Multimedia for Personal Health and Health Care (MMHealth 16'); 10/2016, DOI:10.1145/2985766.2985775.

Yunlong Wang, Le Duan, Simon Butscher, Jens Mueller, Harald Reiterer: "Fingerprints": Detecting Meaningful Moments for Mobile Health Intervention. In Adjunct Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 16' Adjunct); 09/2016, DOI:10.1145/2957265.2965006.


Under Review


Under Revision


CONTENTS

1 GENERAL INTRODUCTION ................................................................. 1
  1.1 MOTIVATION ................................................................................. 2
  1.2 BACKGROUND ............................................................................... 4
  1.3 CONTEXT AND FOCUS ................................................................. 9
  1.4 OBJECTIVES AND DISSERTATION OUTLINE ............................. 9

2 PERSUASIVE TECHNOLOGY FOR REDUCING PROLONGED
SEDENTARY BEHAVIOR ........................................................................ 13
  2.1 ABSTRACT .................................................................................. 14
  2.2 INTRODUCTION ........................................................................... 14
    2.2.1 Effectiveness of Intervention Strategies ............................... 14
    2.2.2 Persuasive Technology .......................................................... 15
    2.2.3 Aim ....................................................................................... 16
  2.3 METHODS .................................................................................... 17
    2.3.1 Data Source ........................................................................... 17
    2.3.2 Study Selection ...................................................................... 17
    2.3.3 Data Coding .......................................................................... 18
  2.4 RESULTS ....................................................................................... 19
    2.4.1 Search and Selection Results ................................................ 19
    2.4.2 Intervention Studies Description ......................................... 20
    2.4.3 Effects of Interventions ........................................................ 21
    2.4.4 Persuasive Technology Analysis .......................................... 21
    2.4.5 In-Depth Analysis of Reminders .......................................... 22
  2.5 PITFALLS ...................................................................................... 27
    2.5.1 PSD Principles Overlapping ................................................ 27
    2.5.2 Behavioral Theories ............................................................... 28
    2.5.3 User Experience ................................................................. 28
  2.6 IMPLICATIONS ............................................................................. 28
  2.7 CONCLUSION .............................................................................. 29

3 INTEGRATING TAXONOMIES INTO THEORY-BASED DIGITAL
HEALTH INTERVENTIONS FOR BEHAVIOR CHANGE ......................... 31
  3.1 ABSTRACT .................................................................................. 32
3.2 INTRODUCTION ........................................................................................................ 32
3.3 RELATED WORK ....................................................................................................... 33
  3.3.1 CeHRes Roadmap ............................................................................................... 33
  3.3.2 Behavioral Intervention Technology Model ...................................................... 34
  3.3.3 IDEAS ................................................................................................................. 34
  3.3.4 The Theory & Techniques Tool .......................................................................... 34
3.4 BEHAVIORAL THEORIES .................................................................................... 35
3.5 DIGITAL HEALTH INTERVENTION TAXONOMY ............................................... 37
3.6 THE HOLISTIC FRAMEWORK .............................................................................. 40
3.7 DISCUSSION ........................................................................................................... 43
3.8 CONCLUSION ......................................................................................................... 43

4 UNDERSTANDING USERS’ MOBILITY PATTERNS: CLUSTERING 45
  4.1 ABSTRACT ............................................................................................................. 46
  4.2 INTRODUCTION .................................................................................................... 46
  4.3 RELATED WORK ................................................................................................... 47
  4.4 APPROACH FOR EXTRACTING POIS ................................................................. 49
    4.4.1 Temporal Information in Human Mobility ..................................................... 49
    4.4.2 Definitions of POIs ......................................................................................... 51
    4.4.3 POI Clustering Algorithm ............................................................................ 52
    4.4.4 Distance Computation in Hierarchical Clustering ....................................... 54
  4.5 EVALUATION ......................................................................................................... 55
    4.5.1 Datasets .......................................................................................................... 55
    4.5.2 Data Pre-processing ......................................................................................... 56
    4.5.3 Compared Approaches and Parameters .................................................... 57
    4.5.4 Evaluation Metrics ........................................................................................ 58
    4.5.5 Results .......................................................................................................... 59
  4.6 DISCUSSION .......................................................................................................... 63
  4.7 CONCLUSION ....................................................................................................... 64

5 UNDERSTANDING USERS’ MOBILITY PATTERNS: PREDICTION 65
  5.1 ABSTRACT ............................................................................................................. 66
  5.2 INTRODUCTION .................................................................................................... 66
  5.3 PROBLEM STATEMENT ....................................................................................... 68
6 SUPPORTING SEDENTARY BEHAVIOR CHANGE BY VISUALIZING PERSONAL MOBILITY PATTERNS AND ACTION PLANNING ON SMARTPHONE .......................... 81

6.1 ABSTRACT ........................................................................................................... 82
6.2 INTRODUCTION .................................................................................................. 82
   6.2.1 Background ..................................................................................................... 82
   6.2.2 Action Planning for Sedentary Behavior Change .............................................. 83
   6.2.3 Visualizations of Mobility Patterns ................................................................. 84
   6.2.4 Study Objectives ............................................................................................ 84
6.3 METHODS ............................................................................................................. 85
   6.3.1 Study Design .................................................................................................. 85
   6.3.2 Participants ..................................................................................................... 86
   6.3.3 SedVis App ..................................................................................................... 86
   6.3.4 Measures ....................................................................................................... 91
   6.3.5 Statistical Analysis ......................................................................................... 92
6.4 RESULTS .............................................................................................................. 93
   6.4.1 Data Collection ............................................................................................... 93
   6.4.2 Participants’ Intention (Control Measure) ...................................................... 93
   6.4.3 RQ1: Effect of Visualization on Participants’ Action Planning....................... 94
8.4.1 Device ................................................................. 119
8.4.2 Task..................................................................... 119
8.4.3 Participants and Procedure................................. 121
8.4.4 Measurements .................................................... 122
8.5 STATISTICAL ANALYSIS ........................................... 123
8.6 RESULTS .................................................................... 124
  8.6.1 Macro-Movements .............................................. 124
  8.6.2 Micro-Movements ................................................ 125
  8.6.3 Head Direction ...................................................... 126
  8.6.4 Workload ............................................................. 127
8.7 DISCUSSION ............................................................... 128
  8.7.1 Explanations of Movements ............................... 128
  8.7.2 Reminders ........................................................... 129
  8.7.3 Limitations and Future Work ............................ 131
8.8 CONCLUSION ............................................................. 131
9 HEALTH BEHAVIOR CHANGE IN HCI: TRENDS, PATTERNS, AND OPPORTUNITIES ................................................................. 133
  9.1 ABSTRACT ................................................................. 134
  9.2 INTRODUCTION ........................................................... 134
  9.3 BACKGROUND .......................................................... 135
    9.3.1 Behavioral Theories ........................................... 135
    9.3.2 Behavior Change Techniques (Strategies) .......... 136
    9.3.3 The Behavioral Intervention Characteristics ........... 137
    9.3.4 The Holistic Framework ...................................... 137
    9.3.5 Related Work .................................................... 137
  9.4 METHODS ................................................................. 138
  9.5 RESULTS AND FINDINGS .......................................... 140
    9.5.1 Trends of Target Behaviors and Paper Types ...... 141
    9.5.2 Patterns in the Selected Intervention Studies ...... 142
    9.5.3 Characteristics of the Selected Intervention Studies ... 147
  9.6 DISCUSSION .............................................................. 149
    9.6.1 Users’ Behavior Priority .................................... 150
9.6.2 Categorizing Target Users ................................................................. 150
9.6.3 Leveraging Users’ Power ................................................................. 151
9.6.4 Longitudinal Study with Game Integration .................................... 151
9.6.5 Cautions for Socialization .............................................................. 152
9.6.6 Embracing AI .................................................................................. 153

9.7 LIMITATIONS .................................................................................. 154
9.8 CONCLUSION .................................................................................... 154

10 GENERAL DISCUSSION ........................................................................ 157

10.1 SUMMARY OF CONTRIBUTIONS .................................................. 158

10.2 IMPLICATIONS AND FUTURE WORK ............................................ 160

10.2.1 The Necessity of More Digital Intervention Studies for Sedentary Behavior Change ........................................................................ 160
10.2.2 Understanding and Leveraging Users’ Context ............................ 161
10.2.3 Enhancing Users’ Interaction and Engagement .......................... 161
10.2.4 Regarding Users’ Health Behavior Change in A Holistic View ....... 162
LIST OF TABLES

TABLE 1.1: RELATED WORK FOUND IN ACM DIGITAL LIBRARY. ........................................ 6
TABLE 1.2: SELECTED WORK OF INTERVENTIONS STUDIES ON SEDENTARY BEHAVIOR
CHANGE. RCT IS SHORT FOR RANDOMIZED CONTROLLED TRAIL. ............................... 7
TABLE 2.1: THE TERMS AND COMBINATION STRATEGY FOR SEARCHING ARTICLES. .... 17
TABLE 2.2: PSD PRINCIPLES. ....................................................................................... 18
TABLE 2.3: INTERVENTION STUDY CHARACTERISTICS. .............................................. 23
TABLE 2.4: THE EFFECTS OF THE INTERVENTION STUDIES. ...................................... 25
TABLE 2.5: REMINDER PROPERTIES IN THE REVIEWED INTERVENTIONS. ............... 27
TABLE 3.1: PSD PRINCIPLES. THE RED PRINCIPLES HAVE COUNTERPARTS WITH THE
SAME OR SIMILAR DEFINITIONS IN THE BCT TAXONOMY. THE BLUE PRINCIPLES
HAVE NO COUNTERPARTS IN THE BCT TAXONOMY BUT CAN ALSO BE REGARDED AS
INTERVENTION STRATEGIES. THE GREEN PRINCIPLES ARE CHARACTERISTICS OF
INTENTION MEDIA............................................................................................................ 39
TABLE 4.1: SUMMARY OF PARAMETERS INFLUENCING POI DEFINITIONS IN THE RELATED
APPROACHES. THE DASHES DEPICT UNUSED PARAMETERS IN THE CORRESPONDING
APPROACHES. CATEGORIES: BLUE: STAY-POINT DETECTION; ORANGE: POI
CLUSTERING; GREEN: DENSITY-BASED CLUSTERING; GRAY: HIERARCHICAL
CLUSTERING ......................................................................................................................... 48
TABLE 4.2: COMPUTATION LOAD COMPARISON BETWEEN HAVERSINE FORMULA AND
THE SIMPLIFIED METHOD. $C^n_2$ REFERS TO THE NUMBER OF COMBINATIONS OF 2
ELEMENTS OUT OF THE TOTAL ELEMENTS N. HERE $C^n_2$ INDICATES THE NUMBER OF
DISTANCES BETWEEN ANY TWO LOCATIONS IN ALL DATA POINTS. ............................... 55
TABLE 4.3: DETAILS ABOUT THE STUDENTLIFE DATASET AND THE MDC DATASET.... 56
TABLE 4.4: THE LIST OF DIFFERENT APPROACHES AND PARAMETERS IN THE
EXPERIMENTS. SP MEANS STAY-POINT DETECTION. THE N IN MINPTS MEANS THE
DATA SIZE. THE OTHER SYMBOLS HAVE THE SAME MEANING AS IN TABLE 4.1. ... 57
TABLE 4.5: THE RESULTS OF THE THIRD EVALUATION ............................................. 63
TABLE 5.1: A SYMBOLIC-TEMPORAL SEQUENCE REPRESENTING HUMAN MOBILITY.... 68
TABLE 5.2: TIME BIN-BASED SYMBOLIC-TEMPORAL SEQUENCE (BASED ON THE SEQUENCE
IN TABLE 5.1), ...................................................................................................................... 69
TABLE 6.1: CODING CRITERIA FOR SPECIFICITY. ..................................................... 91
TABLE 6.2: THE MEASUREMENTS OF THE QUALITY OF THE ACTION PLANS. ............... 94
TABLE 6.3: THE USER EXPERIENCE SCORES BASED ON UEQ. .................................. 97
TABLE 7.1: REASONS FOR NOT TAKING BREAKS WHEN RECEIVING THE PROMPTS IN (LUO
ET AL., 2018). .................................................................................................................. 103
Table 7.2: Related work .......................................................... 104
Table 7.3: The Results of post hoc tests for the durations of prolonged work sessions .......................................................... 109
Table 7.4: The Results of post hoc tests for the number of prolonged work sessions .......................................................... 109
Table 7.5: The participants’ ratings on the reminders, their preferences, and reported reasons. PI refers to the perceived interruption; PU refers to the perceived usefulness. The blue parts indicate the cases when the SedBar was preferred and the assumed attributing ratings in these cases. .......................................................... 112
Table 8.1: Macro-movements and micro-movements in our coding scheme. ........................................................................ 122
Table 8.2: The Bayes factors and the effect sizes (the medians and the credible intervals) of the workload assessment comparison between the two modes. .......................................................... 127
Table 9.1: The query syntax used in ACM digital library .................. 138
Table 9.2: The paper types used in our coding .................................. 139
Table 9.3: The coding items and explanations .................................. 140
Table 9.4: The characteristics used in our coding and the explanations. 141
Table 9.5: The social support types ................................................. 149
LIST OF FIGURES

FIGURE 1.1: AN INTEGRATED AND COMPREHENSIVE VIEW OF ACTIVITY AND INACTIVITY. ADAPTED FROM (AMERICAN THORACIC SOCIETY, 2018)................................. 3

FIGURE 1.2: LINKAGES ACROSS HEALTH SYSTEM CHALLENGES, DIGITAL HEALTH INTERVENTIONS, AND SYSTEM CATEGORIES. ADAPTED FROM (WHO, 2018A)..... 4

FIGURE 1.3: GENERIC DIAGRAM OF THE HEALTH ACTION PROCESS APPROACH. ADAPTED FROM (SCHWARZER, 2008)....................................................... 8

FIGURE 1.4: IDEAS (INTEGRATE, DESIGN, ASSESS, AND SHARE) FRAMEWORK FOR DEVELOPING DIGITAL HEALTH INTERVENTIONS. ADAPTED FROM (SARAH ANN MUMMAH, ROBINSON, KING, GARDNER, & SUTTON, 2016).......................... 8

FIGURE 1.5: THE DIAGRAM SHOWING THE DISSERTATION FOCUS ON THE INTERSECTION OF BEHAVIOR CHANGE FROM PSYCHOLOGY, DATA MINING, AND HCI............... 9

FIGURE 1.6: THE SCHEMATIC OUTLINE OF THIS DISSERTATION........................................ 10

FIGURE 2.1: THE STUDY SELECTION WORKFLOW ........................................................ 20

FIGURE 2.2: PSD PRINCIPLES (SEE DETAILS IN TABLE 2.2) USAGE FREQUENCY ........... 22


FIGURE 3.2: THE TRANSTHEORETICAL MODEL OF CHANGE (TTM), ADAPTED FROM (AMERICAN SOCIETY ON AGING AND AMERICAN SOCIETY OF CONSULTANT PHARMACISTS FOUNDATION;, 2012). THIS MODEL DIVIDES THE BEHAVIOR CHANGE PROCESS INTO FIVE STAGES, NAMELY PRECONTEMPLATION, CONTEMPLATION, PREPARATION, ACTION, AND MAINTENANCE AND RELAPSE PREVENTION. DEPENDING ON THE STAGE OF CHANGE, DIFFERENT STRATEGIES COULD BE APPLIED ACCORDINGLY TO MAKE THE INTERVENTION EFFECTIVE. ..... 36

FIGURE 3.3: THE WORD CLOUD OF BCTS USED IN THE STUDY COLLECTION OF 405 INTERVENTION STUDIES FROM THE OFFICIAL WEBSITE OF THE BCT TAXONOMY (BCT TAXONOMY RESEARCH TEAM, 2014) ....................................................... 38

FIGURE 3.4: THE DIAGRAM OF OUR DHI TAXONOMY. THE BLUE PART IS THE STRATEGY PART, WHILE THE GREEN PART IS THE CHARACTERISTICS PART............. 40

FIGURE 3.5: THE DIAGRAM OF TUDER (TARGETING, UNDERSTANDING, DESIGNING, EVALUATING AND REFINING)................................................................. 41

FIGURE 3.6: THE CHECKLIST FOR USING TUDER......................................................... 43

FIGURE 4.1: THE WORKFLOW ADOPTED IN RELATED WORK (ikanovic & mollgaard, 2017; montoliu et al., 2013; ye et al., 2009)......................................................... 48
FIGURE 4.2: THE WORKFLOW OF OUR APPROACH. IT CONTAINS THREE STEPS: DATA PRE-
PROCESSING, GLOBAL POIs CLUSTERING, AND LOCAL POIs CLUSTERING...........49

FIGURE 4.3: LOCATION DATA POINTS AND TRANSITIONS OF TWO COLLEGE STUDENTS.
IN THE 2D MAP (TOP) THE BLACK DOTS REPRESENT DATA POINTS OF LOCATION,
AND THE LINES SHOW THE TRANSITIONS. THE BLACK CIRCLES DEMONSTRATE TWO
SPATIALLY DISTINCT CLUSTERS WHILE THE BLACK RECTANGLE HIGHLIGHTS A
POTENTIAL CLUSTER THAT CANNOT BE RECOGNIZED USING ONLY SPATIAL
INFORMATION. BY ADDING THE TIME DIMENSION, THE BLACK RECTANGLE IN THE
3D FIGURE (BOTTOM) SHOWS THE SIGNIFICANT VISIT FREQUENCY OF THE
LOCATION IN THE BLACK RECTANGLE IN THE 2D MAP..........................50

FIGURE 4.4: THE RELATIONSHIP BETWEEN GLOBAL POIs AND LOCAL POIs. ........51

FIGURE 4.5: ASSUME THE MINIMUM DISTANCE BETWEEN THE TWO POTENTIAL POIs IN
TWO BUILDINGS IS 45 METERS, THEN A STAY-POINT DETECTION OR DBSCAN WITH
A DISTANCE THRESHOLD OF 50 METERS CANNOT DIVIDE THE TWO CLUSTERS. THE
GREEN COLOR INDICATES LOCATIONS BELONG TO ONE POI RESULTED FROM
DBSCAN. ...............................................................................................52

FIGURE 4.6: THE EVALUATION RESULTS OF DIFFERENT APPROACHES FOR THE
STUDENTLIFE DATASET AND THE MDC DATASET. THE TWO LEFTMOST GRAPHS
SHOW THE NUMBERS OF EXTRACTED POIs; THE TWO MIDDLE GRAPHS PRESENT THE
PREDICTABILITY LIMIT OF THE POI SEQUENCES; THE TWO RIGHTMOST GRAPHS
REPRESENT THE COMPUTATION TIME. ....................................................61

FIGURE 4.7: THE RESULTS OF THE SECOND EVALUATION. THE TWO LEFTMOST GRAPHS
SHOW THE NUMBERS OF EXTRACTED POIs; THE TWO MIDDLE GRAPHS PRESENT THE
PREDICTABILITY LIMIT OF THE POI SEQUENCES; THE TWO RIGHTMOST GRAPHS
REPRESENT THE COMPUTATION TIME. ....................................................62

FIGURE 4.8: THE MULTIPLE POIs CAUSED BY THE NOISE IN GPS DATA..............63

FIGURE 5.1: THE HUMAN MOBILITY MODEL, CONTAINING CONCRETE PERSON’S
LOCATIONS (CONTINUOUS BOXES) AND PREDICTED ONES (DASHED BOXES). ......69

FIGURE 5.2: THE FIVE SELECTED METHODS IN OUR EVALUATION. .....................71

FIGURE 5.3: DISTRIBUTION OF SEQUENCE LENGTH AND POI NUMBER OF THE PREPARED
DATA FROM THE STUDENTLIFE DATASET AND MDC DATASET ..................73

FIGURE 5.4: THE RESULTS OF INCREMENTAL-PREDICTION ACCURACY (IPA) FOR USER #2
FROM STUDENTLIFE DATASET (LEFT) AND USER #8 FROM MDC DATASET (RIGHT).
................................................................................................................75

FIGURE 5.5: THE AVERAGE ACCURACY AND TIMES OF WINNING (ToW) OF EACH
METHOD FOR THE STUDENTLIFE AND MDC DATASETS. ALZ PRESENTS ACTIVE
LEZI AND PL DENOTES PREDICTABILITY LIMIT........................................76

FIGURE 5.6: USER’S PATTERN AND POI PREDICTION VISUALIZER (UP3V). THE MAP
(LEFT SIDE) DISPLAYS THE PLACES/SPATIAL INFORMATION AND AN INFORMATION
window (right side) shows the results per selected user. The information panel shows for the selected POI: (1) the prediction accuracy, (2) the arrival time distribution per weekday, and (3) the stay periods. ................................................................. 78

Figure 6.1: The mixed design of our three-week study ........................................... 85
Figure 6.2: The always-on notification of SedVis on a user’s smartphone.. 87
Figure 6.3: The dashboard of SedVis................................................................. 88
Figure 6.4: The mobility patterns in the daily visualization mode ................. 89
Figure 6.5: The multi-day visualization ............................................................. 90
Figure 6.6: The action planning function in the app. ........................................... 90
Figure 6.7: A graphical representation of a Bayes factor classification and the interpretation, adapted from (Doorn et al., 2019) ................. 93
Figure 6.8: The participants’ daily sedentary hours based on the app-logged data during the baseline week and the intervention weeks. The bars refer to the confidence intervals with 95% confidence level ....... 95
Figure 6.9: The daily frequency of participants checking the visualizations in SedVis through the notification and the home screen ....................... 96
Figure 6.10: The scatter plot of the participants’ daily frequency of checking the visualizations and their change in daily sedentary hours in Group A. The change of daily sedentary hours (x-axis) equals to the daily sedentary hours during the intervention weeks minus the counterparts during the baseline week. Thus, negative values indicate a reduction in sedentary behavior ......................................................... 97

Figure 7.1: The state transitions between working and breaking. W-B condition: a user is absent for more than a minute in the working state. B-W condition: a user starts to use the mouse or the keyboard in the breaking state ................................................................. 105

Figure 7.2: The prompt interface. In our study, it shows up when the current duration of working state exceeds the pre-defined duration. It disappears when the user clicks on the button or after 30 seconds without interaction ................................................................. 106

Figure 7.3: The SedBar. In the working state, the progress bar color is orange by default, while the progress indicates the working duration. In the breaking state, the progress bar color is blue by default, while the progress loops like a battery charging animation ......................... 106

Figure 7.4: The study procedure. The dots represent the appointments during the study. The SedBar and the Prompt use the same context detection method and follow the state transition conditions as shown in Figure 7.1. The dots refer to the appointments ........................................... 107
Figure 7.5: The results of post hoc tests for the prolonged work sessions (t > 30 minutes) in three conditions for each participant. The horizontal axis represents duration in minutes. The red bar indicates the statistically significant difference compared to the selected condition (the blue bar). ................................................................. 108

Figure 7.6: The results of post hoc tests for the prolonged work sessions (t > 30 minutes) in three conditions aggregated of all participants. The horizontal axis represents duration in minutes. The red bar indicates the statistically significant difference compared to the selected condition (the blue bar). ................................................................. 109

Figure 7.7: The results of post hoc tests for the numbers of the prolonged work session (t > 30 minutes) in three conditions aggregated of all participants. The horizontal axis represents duration in minutes. ... 110

Figure 7.8: The bar chart (with the error bar of standard deviation) of each participant’s perceived interruption for SedBar and the prompt. The smaller the better. The solid blue bars highlight the cases where the SedBar was perceived less interruptive than the prompt. ....................... 110

Figure 7.9: The bar chart (with the error bar of standard deviation) of each participant’s perceived usefulness for SedBar and the prompt. The larger the better. The solid blue bars highlight the cases where the SedBar was perceived more useful than the prompt. ....................... 111

Figure 7.10: The bar chart (with Confidence Interval) of the perceived interruption and usefulness aggregated from all the participants in the two intervention conditions. ................................................................. 111

Figure 8.1: An question example in the trivial quiz................................................. 120

Figure 8.2: The physical environment in the study................................................. 120

Figure 8.3: A graphical representation of a Bayes factor classification and the interpretation, adapted from (Doorn et al., 2019)......................... 123

Figure 8.4: The boxplot of macro-movements in the spatial-mapping (SM) mode, and the tag-along (TA) mode. The boxplot of within-subject difference (TA-SM) is to show the effect intuitively. The circles in the boxplots refer to the means................................................................. 124

Figure 8.5: The boxplot of sit-stand transitions and body direction/position changes in the spatial-mapping (SM) mode, the tag-along (TA) mode, and the within-subject difference (TA-SM). ........................................ 125

Figure 8.6: The boxplot of users’ standing duration in the spatial-mapping (SM) mode, the tag-along (TA) mode, and the within-subject difference (TA-SM). The cross in TA-SM refers to the outlier. ......................... 125
Figure 8.7: The boxplot of the numbers of adjusting HoloLens and head movements in the spatial-mapping (SM) mode, the tag-along (TA) mode, and the within-subject difference (TA-SM). ................................................................. 126

Figure 8.8: The boxplot of the numbers of adjusting HoloLens and head movements in the spatial-mapping (SM) mode, the tag-along (TA) mode, and the within-subject difference (TA-SM). ................................................................. 127

Figure 8.9: The bar chart of the pitch angle of HoloLens during the study in the spatial-mapping (SM) mode, and the tag-along (TA) mode. The data from participants #5 and #6 contribute to the most difference on average. ........................................................................................................... 127

Figure 8.10: The boxplots of the participants’ scores of the six workload factors in the NASA-TLX questionnaire. ................................................................................................................................. 128

Figure 8.11: Each row represents a one-hour study session. The green part refers to standing; the red part refers to sitting; the white part means the participant stopped because of eye fatigue................................. 130

Figure 8.12: Selected bad postures of the participants during our study. 130

Figure 9.1: The paper frequency distribution in the field of health behavior change in the HCI community. Note that we extracted the original data from the ACM digital library on August 23, 2018, when some conferences for this year have not taken place. The majority of the papers are from conference proceedings, while a small part of them are from journals (e.g., Personal and Ubiquitous Computing or PUC in the figure). ................................................................................................................................. 135

Figure 9.2: The word cloud of behavior change strategies coded from the papers listed on the official website of BCT taxonomy (N=405). The bigger the font is, the more frequently the strategy was used. ........ 136

Figure 9.3: The workflow of screening and selecting papers. 139

Figure 9.4: The target behaviors in the reviewed papers over time (N=648). ................................................................. 141

Figure 9.5: The paper category in the reviewed papers over time (N=648)... 142

Figure 9.6: The percentages for the used measurements in the reviewed studies. ................................................................................................................................. 142

Figure 9.7: The distribution of the target user group ................................................................. 143

Figure 9.8: The distribution of the used behavioral theories. SRT refers to the self-regulation theory; HBM refers to the health belief model; IBM refers to the integrated behavioral model. The other acronyms can be found in the following content ................................................................. 144

Figure 9.9: Behavioral theories and the ways that they were used in the reviewed studies. I – informing design, G – guide evaluation strategies,
S – SELECTING TARGET USERS. NOTE THAT, IN THIS ALLUVIAL GRAPH, THE RELATIVE SIZES OF THE BARS FOR EACH BEHAVIORAL THEORY DO NOT EXACTLY REPRESENT THEIR USE FREQUENCY. IN ONE STUDY, A BEHAVIORAL THEORY CAN BE USED FOR BOTH INFORMING DESIGN AND GUIDING EVALUATION STRATEGIES. 145


FIGURE 9.11: THE TOP-4 TARGET BEHAVIORS AND THE CORRESPONDING BEHAVIOR CHANGE STRATEGIES GROUPS IN THE REVIEWED STUDIES. NOTE THAT, IN THIS ALLUVIAL GRAPH, THE RELATIVE SIZES OF THE BARS FOR EACH TARGET BEHAVIOR DO NOT EXACTLY REPRESENT THEIR FREQUENCY. GIVEN THE TARGET BEHAVIOR IN ONE STUDY, SEVERAL STRATEGIES MIGHT BE USED. .................................................... 146


FIGURE 9.14: THE STUDY DURATION OF THE STUDIES WITH GAME INTEGRATION...... 152

FIGURE 9.15: THE DISTRIBUTION OF THE REVIEWED STUDY DURATIONS.................. 152
LIST OF ABBREVIATIONS AND ACRONYMS

Association for Computing Machinery (ACM)
Augmented Reality (AR)
Behavior Change Techniques (BCTs)
Behavioral Intervention Technology (BIT)
Confidence Interval (CI)
Digital Health Interventions (DHIs)
Electronic Health (eHealth)
Head-Mounted Display (HMD)
Human-Computer Interaction (HCI)
Information and Communication Technologies (ICT)
Metabolic Equivalent (MET)
Mobile Health (mHealth)
Moderate- and Vigorous-Intensity Physical Activity (MVPA)
Non-Communicable Diseases (NCDs)
Personal Computer (PC)
Persuasive System Design (PSD)
Place of Interest (POI)
Sedentary Behavior (SB)
Short Message Service (SMS)
Traditional Chinese Medicine (TCM)
World Health Organization (WHO)
Chapter 1: General Introduction

1 GENERAL INTRODUCTION

“上医治未病。"

The best doctor prevents you from diseases.

- 《黄帝内经》
  Esoteric Scripture of the Yellow Emperor
  (A TCM classic)
1.1 Motivation

According to a report from the World Health Organization (WHO) in 2018 (WHO, 2018c), non-communicable diseases (NCDs) are the leading cause of death globally, which were responsible for 71% (41 million) of the 57 million deaths occurred globally during the year of 2016. The major NCDs include cardiovascular diseases, cancers, chronic respiratory and diabetes, which are causally linked with four leading behavioral risk factors: tobacco use, harmful use of alcohol, physical inactivity, and unhealthy diet. Physiologically, these unhealthy behaviors could raise people’s blood pressure, blood glucose, blood lipids, and body weight. For some chronic diseases (e.g., type 2 diabetes), the current medicine could only control their development in patients but not get rid of them. For many kinds of cancers, especially in the later stage, patients could get very limited help from the hospital. Therefore, improving our daily behaviors and developing healthy lifestyles are the best way, sometimes the only way, to live longer and better.

Physical inactivity becomes prevalent along with the industry development: machines and computers have increasingly replaced occupational physical-activities; modern office work requires people to sit in front of their computers all day long. In 2016, there were 42.2% (35.5-49.2) of the adult in Germany did not meet current WHO recommendations on physical activity for health (Guthold, Stevens, Riley, & Bull, 2018), which require any of the following criteria:

- At least 150 minutes of moderate-intensity physical activity per week;
- At least 75 minutes of vigorous-intensity physical activity per week;
- A combination of moderate- and vigorous-intensity physical activity (MVPA) per week, accumulating at least 600 metabolic equivalents (METs)-minutes per week.

Meeting physical activity recommendations could largely decrease the risk of many diseases, but it might be still not enough. Evidence has shown that too much sitting and too little MVPA represent separate and distinct risk factors for many non-communicable diseases (e.g., cardiovascular disease, diabetes, cancer) (Knaeps et al., 2016; Shrestha et al., 2016). In other words, only increasing physical activity cannot completely attenuate the negative health consequences of too much sitting (prolonged sedentary behavior).

Sedentary behavior (SB) refers to any waking behavior characterized by an energy expenditure less than or equal to 1.5 METs1 (or metabolic equivalents) with a sitting or reclining posture during tasks such as working at a desk and watching TV (SBRN, 2012). Figure 1.1 illustrates the definitions of sedentary behavior and physical activities according to metabolic equivalents. Prolonged

---

1 One MET is roughly equal to the energy expended when a person is sitting quietly.
sedentary behavior has been defined as maintaining sedentary behavior for more than 30 minutes, and this definition has been well adopted in this domain (Hadgraft et al., 2016; Henson et al., 2016). “Sitting has become the smoking of our generation (Merchant, 2013).” It has become a common sense that prolonged sedentary behavior is very unhealthy.

**Figure 1.1: An integrated and comprehensive view of activity and inactivity. Adapted from (American Thoracic Society, 2018).**

Although the mechanism of the detrimental effect caused by prolonged sedentary behavior is not well understood, several studies have provided clues. A study (D. Dunstan, Kingwell, Larsen, & Healy, 2012) with 19 participants showed that interrupting sitting time with short bouts (2-min bouts every 20-min sitting) of light- or moderate-intensity walking could both lower postprandial glucose and insulin levels in overweight/obese adults. Another study (Rodriguez-Hernandez, Martin, Pascoe, Roberts, & Wadsworth, 2018) with ten obese women found that 5-min bouts of moderate-intensity walking breaks after every 30-min sitting attenuated 2-hour postprandial glucose excursions, while 2-min bouts of that did not. In addition to postprandial glucose and insulin levels, another study (Bergouignan et al., 2016) with 30 sedentary adults also found other benefits of frequent breaks for sedentary people. Frequent short breaking (5-min bouts of moderate-intensity treadmill walking every hour) improved mood, decreased levels of fatigue, and reduced food cravings at the end of the day compared to the condition of 6-h uninterrupted sitting; the condition of one 30-min moderate-intensity treadmill walking in the morning did not have such effects.

Although the dose-response and frequency-response effects of breaking sedentary behavior have not been consistently and systematically evaluated, we believe that breaking the long sedentary sessions of office workers as frequently as acceptable could improve their occupational health. Therefore, in the dissertation, we aim to explore the potential of technologies to help the sedentary population reduce their sedentary behavior.
1.2 Background

The prevalence of sitting opportunities in our daily life leads to high habit strength of sedentary behavior, which makes it difficult to change in the long term (D. W. Dunstan, Healy, Sugiyama, & Owen, 2010). Chu et al. (Chu et al., 2016) divided intervention strategies of reducing sedentary behavior to three categories in their review: (1) educational/behavioral (e.g., goal setting, action planning, and self-monitoring); (2) environmental changes (e.g., sit-stand workstation and treadmill desk); (3) multi-component (e.g., sit-stand workstation plus goal-setting). The environmental changes might require policy support and additional facilities, which hinder its application in large scale. Therefore, low-cost and light-weight solutions are needed. To this end, we focus on developing software on smartphones and personal computers (PCs) to provide digital health interventions (DHIs) (WHO, 2018a) for sedentary behavior change.

1.2 Background

DHIs should be developed to tackle the challenges of healthcare management and health behavior change, as shown in Figure 1.2. For sedentary behavior change, users’ digital devices could be used to address the lack of access to information or data and the loss to follow-up of clients (users). This potential of digital devices is attributed to their prevalence in our daily life: smartphones know our mobility as we always take them; PCs are aware of our work durations because we depend on them for office work.

Figure 1.2: Linkages across Health System Challenges, Digital Health Interventions, and System Categories. Adapted from (WHO, 2018a).
Before we started to design our digital health interventions for sedentary behavior change, we first reviewed the existing related work to answer three basic questions:

- What have digital technologies been created and used for sedentary behavior change?
- What is the effect of existing DHIs for office workers’ sedentary behavior change?
- What is the mechanism underlying behavior change, and how can we design DHIs accordingly?

We address the first question by searching and reviewing the related work in ACM digital library, as shown in Table 1.1. We found that all the reviewed work aimed to develop context-aware reminders for reducing sedentary behavior, where the context included the inactivity/sitting duration and the sitting posture. Smartphones, dedicated sit pats, motion trackers, cameras, Kinect, and keyboard/mouse were used as the context detection tools; mobile apps (including visualizations, notifications, vibration), SMS, PC prompts, dedicated light tube, and beamer-projected figures were used as the media to deliver reminders. However, most of the reviewed digital technologies were concepts, proof-of-concept, or feasibility studies (without statistical analysis). Only two study (Grundgeiger et al., 2017; van Dantzig, Geleijnse, & van Halteren, 2013) showed statistical results of the intervention effect, which is due to the fact that the computer science community – the human-computer interaction community, to be exact – concerns more about the technology part than the intervention part in health behavior change systems.

To answer the second question, we looked into more extensive literature databases covering Google Scholar, ACM digital library, JMIR2, and PubMed3. Table 1.2 listed the eight selected compared intervention studies, which we will provide a detailed analysis in Chapter 2. Compared to the work in Table 1.1, these eight studies adopted more serious study design (RCTs or quasi-experiments). Most studies used validated motion trackers to access participants’ sedentary behavior. Based on these studies, we could answer our second question: (1) the hourly PC prompt was the most frequently used intervention method; (2) the hourly PC prompt alone had no significant effect on reducing sedentary behavior at work; (3) the hourly PC prompt coupling with educational/informative session was effective.

---

2 http://www.jmir.org/
### Table 1.1: Related work found in ACM digital library.

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Software</th>
<th>Study Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirano et al.</td>
<td>Smartphone; Bar chart showing bouts of steps across the day in an app; progress bar showing the progress to the daily steps goal in app-widget; reminding every 30-min inactivity (vibration).</td>
<td>A feasibility study (N=8, four weeks)</td>
</tr>
<tr>
<td>Wang and Yu</td>
<td>Sit pad; Message providing warning, and activity/relaxation advice.</td>
<td>Proof-of-concept</td>
</tr>
<tr>
<td>Dantzig et al.</td>
<td>Smartphone; motion tracker; PC; Reminding every 60-min inactivity (smartphone prompt); reminding every 30-min inactivity (smartphone SMS).</td>
<td>A feasibility study (N=8, one day); a compared study (86 participants, seven weeks)</td>
</tr>
<tr>
<td>Mateevitsi et al.</td>
<td>Infrared motion sensor; light tube; Changing the color of a light tube (in front of the PC) when sitting and pulsing every 45-min uninterrupted sitting.</td>
<td>A feasibility study (N=8, five days)</td>
</tr>
<tr>
<td>Ferreira et al.</td>
<td>PC with webcam; Changing PC wallpapers; prompts; changing the posture of an origami construction.</td>
<td>Concept</td>
</tr>
<tr>
<td>Min et al.</td>
<td>Sit pad; smartphone; Reminding every 30-min unbalanced sitting or every 60-min balanced sitting to react to the virtual pet in a mobile app.</td>
<td>Proof-of-concept</td>
</tr>
<tr>
<td>Van Schagen et al.</td>
<td>Smartphone; PC; Reminding every 60-min sitting to react to a mobile game.</td>
<td>Concept</td>
</tr>
<tr>
<td>Pinder et al.</td>
<td>Smartphone; Reminding every time when users unlock their smartphones (text – “active :)”).</td>
<td>Concept</td>
</tr>
<tr>
<td>Grundgeiger et al.</td>
<td>Smartphone; Reminding every 30-min inactivity (smartphone vibration).</td>
<td>A pilot field study (N=5, five days)</td>
</tr>
<tr>
<td>Wölfel</td>
<td>Kinect; Changing the state of a beamer-projected orchid based on users’ postures and sitting time.</td>
<td>A feasibility lab study (N=16, 3 hours)</td>
</tr>
<tr>
<td>Luo et al.</td>
<td>PC; Reminding every 60-min inactivity (PC prompt).</td>
<td>An exploratory field study (N=25, three weeks)</td>
</tr>
</tbody>
</table>
Table 1.2: Selected work of interventions studies on sedentary behavior change. RCT is short for randomized controlled trial.

<table>
<thead>
<tr>
<th>Study</th>
<th>Hardware</th>
<th>Software</th>
<th>Study Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evans et al. (2012)</td>
<td>PC; motion tracker (activPAL)</td>
<td>Reminding every 30-min (PC prompt).</td>
<td>RCT (N=28, 10 days)</td>
</tr>
<tr>
<td>Donath et al. (2015)</td>
<td>PC with height-adjustable desk; motion tracker (ActiGraph)</td>
<td>Reminding three times per day (PC prompt).</td>
<td>RCT (N=31, 13 weeks)</td>
</tr>
<tr>
<td>Puig-Ribera et al. (2015)</td>
<td>PC/smartphone; motion tracker</td>
<td>Reminding every seven or four days (email).</td>
<td>Quasi-experiment (N=190, 22 weeks)</td>
</tr>
<tr>
<td>Brakenridge et al. (2016)</td>
<td>Smartphone; motion tracker (LUMOback and ActivPAL)</td>
<td>Real-time feedback and prompts without mentioning frequency (mobile app).</td>
<td>RCT (N=153, 12 weeks)</td>
</tr>
<tr>
<td>Taylor et al. (2016)</td>
<td>PC</td>
<td>Reminding five hourly times per day (PC prompt).</td>
<td>RCT (N=106, 6 months)</td>
</tr>
<tr>
<td>De Cocker et al. (2016)</td>
<td>PC; motion tracker (activPAL)</td>
<td>Web-based computer-tailored suggestions.</td>
<td>RCT (N=93, 3 months)</td>
</tr>
<tr>
<td>Gilson et al. (2016)</td>
<td>PC; sit pad; motion tracker (GENEActiv)</td>
<td>Reminding 30-min sitting (PC prompt); changing the prompt color after 30-min and then 60-min sitting.</td>
<td>Quasi-experiment (N=57, 5 months)</td>
</tr>
<tr>
<td>Urda et al. (2016)</td>
<td>PC; motion tracker (ActivPAL)</td>
<td>Reminding every 60-min (PC prompt).</td>
<td>RCT (N=44, 2 weeks)</td>
</tr>
</tbody>
</table>

To address the third question, we need to step out the scope of sedentary behavior and overview the whole picture of human behavior. In psychology, two groups of theories have been used to explain and predict human behavior: continuum theories and stage theories. Continuum theories assume people’s behaviors are caused by a set of factors, which can be represented by a path model indicating the causal relationship. As shown in Figure 1.3, for example, the intention is caused by action self-efficacy, outcome expectancies, and risk perception. By contrast, stage theories (e.g., the transtheoretical model (Prochaska & Velicer, 1997)) assume people change their behavior in a process including several stages. The factors pushing people from one stage to the next are different, which lead to strategies for behavior change interventions, respectively. It has shown that theory-based behavior change interventions are more effective than others (R. Davis, Campbell, Hildon, Hobbs, & Michie, 2015; Glanz & Bishop, 2010). Nevertheless, behavioral theories could also be ignored.
1.2 Background

(Y. Wang, Wu, Lange, Fadhill, & Reiterer, 2018) or misused (Hekler, Klasnja, Froehlich, & Buman, 2013). Although behavioral theories allow to explain and predict behavior, they lack the guidance of translating into operational techniques.

Figure 1.3: Generic diagram of the Health Action Process Approach. Adapted from (Schwarzer, 2008).

From our reviewed studies, we found several existing frameworks to guide the DHI development. As an example, Figure 1.4 shows the IDEAS framework for developing digital health interventions, which consists of ten phases grouped into four categories (integrate, design, access, and share). The key research interest in DHI research is to find out what intervention strategies/elements are effective in specific behavior change tasks. Therefore, an insistent and comprehensive taxonomy is necessary. The IDEAS framework mentioned the well-known taxonomy of behavior change techniques (BCTs) (Michie et al., 2013) and emphasized the process motivators (Robinson, 2010) as intervention strategies. However, the more popular taxonomy in the HCI community is the framework of persuasive technology (Fogg, 2003) or persuasive system design (PSD) principles (Oinas-kukkonen & Harjumaa, 2009). This difference intrigues us to develop a holistic framework integrating the taxonomies, especially from the perspective of HCI. We will illustrate our framework in Chapter 3.

Figure 1.4: IDEAS (Integrate, Design, Assess, and Share) framework for developing digital health interventions. Adapted from (Sarah Ann Mummah, Robinson, King, Gardner, & Sutton, 2016).
1.3 Context and Focus

Answering the three questions in the previous section is to prepare the research of DHIs for sedentary behavior change. The main focus of this dissertation is to conduct empirical studies of designing and evaluating applications on smartphones and PCs for reducing the sedentary behavior of office workers and college students. The reason for using applications on smartphones and PCs is quite straightforward: they are the most convenient devices enabling us to both collect the data of users’ sedentary behavior and deliver interventions to users. By mining the collected user behavior data, we could develop context-aware and personalized interventions. Extra devices - like wearables and sit pads – will hinder the large-scale sharing of our technologies to help more sedentary people. The reason for targeting office workers and college students is that they are both vulnerable groups of the sedentary lifestyle and available subjects to us. Through the evaluation of our applications based on behavior change theories, we aim to (1) understand and explain the process of users’ sedentary behavior change and (2) provide implications for DHI development for sedentary behavior change. Besides, we also (3) explore the potential impact of augmented-reality head-mounted devices (e.g., HoloLens) on office workers’ sedentary behavior. Therefore, the focus of this dissertation stands at the intersection of behavior change from psychology, data mining, and HCI, as shown in Figure 1.5.

![Figure 1.5: The diagram showing the dissertation focus on the intersection of behavior change from psychology, data mining, and HCI.](image)

1.4 Objectives and Dissertation Outline

With the focus presented in the previous section, we have multiple research objectives in this dissertation. Accordingly, we arrange the remainder of this dissertation following the structure, as shown in Figure 1.6.
Chapter 2 presents the systematic review of persuasive technologies for sedentary behavior change, which is to reveal the applied digital technologies in sedentary behavior change interventions for office workers and their effect among critically evaluated studies. We firstly introduce the related reviews and the framework we use for coding the reviewed studies. Then we describe the review method and criteria. Following, we provide the review results on the selected studies, the intervention effects, and the used persuasive technologies (especially the most used PC reminders). We also discussed the pitfalls found during our review process and implications for related work. Chapter 2 serves as the research background and start point for our research on digital health interventions for sedentary behavior change.

Chapter 3 illustrates the holistic framework we proposed for designing and reporting DHIs. We firstly talk about the motivation of proposing this framework, which is to enable a better quality of evidence accumulation in DHI research. Then we discuss the advantages and disadvantages of three representative related work, based on which we further review behavioral theories and propose our DHI taxonomy. Finally, we present the holistic framework through each part and discuss its use cases and limitations. Chapter 3 provides the basic methodology for our research on digital health interventions for sedentary behavior change.

Chapter 4 describes a clustering approach to extracting places of interest (POIs) from the spatio-temporal data of users’ mobility, which is used in our mobile app (SedVis, see details in Chapter 6) to find out users’ sedentary places and the corresponding temporal patterns. In the beginning, we introduce the problem of POI clustering and the drawbacks of existing approaches. Then we explain why the temporal information is important in POI clustering of human mobility data, which is not well-used in prior work. After demonstrating how temporal
constraints in human mobility data could be used, we propose our POI clustering approach. We then report the evaluation results of the proposed approach using several metrics based on two datasets, followed by the discussion of its limitations.

Chapter 5 evaluates a group of next-place prediction methods using the sequential data of places clustered from spatio-temporal data of human mobility. The objective of this chapter is to understand how to effectively use the embedded sequential and temporal patterns in human mobility data for the problem of next-place prediction, which could be potentially used for context-aware DHIs. We discuss the related work, explain the used patterns and compared methods, and evaluate them using our proposed metrics based on two datasets. We also demonstrate our visualization tool for individual data analysis. Finally, we discuss our findings and limitations. Due to the large predictability diversity of individuals’ mobility based on the smartphone-logged data, we did not apply next-place prediction in our interventions studies.

Chapter 6 presents a DHI study for sedentary behavior change using a mobile app (SedVis), which aims to evaluate the effects of mobility pattern visualizations on action planning and their effects on reducing sedentary behavior. We firstly introduce the related work of action planning for sedentary behavior change and visualizations of mobility patterns. Then we describe our study method and our dedicated mobile app. Following we report our study results, and we discuss our findings, implications for future work, and limitations in the end.

Chapter 7 presents another DHI study for sedentary behavior change using a PC application (SedBar), which aims to compare the effects of two context-aware reminders on reducing sedentary behavior. We firstly introduce the related work of reminders for sedentary behavior change. Then we describe our context-aware reminding system and propose a reminder visualization to compare with the prompt. Following we report our study design and results. In the end, we discuss the findings, implications, and limitations.

Chapter 8 shows an exploratory study (SedHolo) of the potential impact of augmented-reality head-mounted displays (AR-HMDs) on the movement behavior at office work. We firstly talk about the health impact of the fixed-screen style of office work and our hypotheses of the potential impact of AR-HMDs on the movement behavior with different level of freedom and flexibility of moving the virtual screen. Then we show the related work on ergonomics with digital technologies, specifically their impact on the movement behavior. Following we present our study design and results. Finally, we discuss our findings, limitations, and future work.

Chapter 9 steps out of the scope of sedentary behavior change and presents a systematic review of the whole health behavior change research in the HCI community, which aims to understand the research trend, reveal the research patterns, and discover research opportunities.
Chapter 10 summarizes the contributions, integrates the implications from our intervention studies, discusses the limitations of the dissertation, and points out the directions of future work.
2 Persuasive Technology for Reducing Prolonged Sedentary Behavior

久视伤血，久卧伤气，久坐伤肉，久立伤骨，久行伤筋。

Prolonged watching, lying, sitting, standing, and walking are all detrimental for health, respectively.

- 《黄帝内经》
  *Esoteric Scripture of the Yellow Emperor*
  *(A TCM Classic)*


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Lingdan Wu, Jan-Philipp Lange, and Ahmed Fadhil helped in conducting the study, analyzing parts of the study data, and reviewing the paper. Harald Reiterer supervised the work and reviewed this paper.
2.1 Abstract

In this chapter, we use the framework of persuasive system design (PSD) principles to investigate the utilization and effectiveness of persuasive technology in intervention studies at reducing sedentary behavior at work. This systematic review reveals that reminders are the most frequently used PSD principle. The analysis on reminders shows that hourly PC reminders alone have no significant effect on reducing sedentary behavior at work, while coupling with education or other informative session seems to be promising. Details of deployed persuasive technology with behavioral theories and user experience evaluation are lacking and expected to be reported explicitly in the future intervention studies.

2.2 Introduction

2.2.1 Effectiveness of Intervention Strategies

To reduce sedentary behavior and make people be more active, many intervention studies have been conducted and several review papers examining the effectiveness of the interventions are available. Nevertheless, none of them focused on the impact of technology on reducing SB.

A pioneering work by Chau and colleagues (Chau et al., 2010) addressed the effect of workplace interventions in reducing SB. It reviewed studies on the effectiveness of workplace interventions to reduce sitting behavior. However, all the selected studies focused on increasing physical activity with reducing sitting time as a side effect or a secondary goal, which may weaken the effectiveness of the interventions on reducing sedentary behavior (Gardner, Smith, Lorencatto, Hamer, & Biddle, 2016; Prince, Saunders, Gresty, & Reid, 2014).

Based on 11 studies that aimed to improve health conditions at work by reducing SB, a narrative review by Healy et al. (Healy et al., 2012) supported the use of multiple strategies (e.g., increasing the number of breaks from sitting time, changing postural, focusing on ergonomic changes to the individual workspace, altering the built design of the broader workplace, and using multiple strategies) to reduce prolonged workplace sitting, since these strategies could not only improve the participants’ health conditions at work, but also typically had a beneficial or neutral impact on productivity, absenteeism and injury costs.

With a meta-analysis, Shrestha et al. (Shrestha, Ijaz, Kt, Kumar, & Cp, 2016) aimed to provide a more precise understanding of the workplace interventions for reducing the SB at work. Eight studies were included with a total of 1125 participants who were grouped into three categories: physical workplace changes, policy changes, and information and counseling. The results showed that sit-stand desks can reduce sitting time at work, while the effects of policy changes and information and counseling are inconsistent. It was pointed out that all the reviewed studies were at high risk of bias and the quality of the evidence
was low due to small sample sizes and poor research design (i.e., inadequate randomization, allocation concealment, blinding of outcome assessment, incomplete outcome data, or selective reporting).

The problem with the low quality of evaluation methods was also mentioned in the review by Gardner and colleagues. In this review, Gardner et al. (Gardner et al., 2016) focused on identifying effective behavior change strategies used in sedentary behavior reduction interventions, based on intervention functions (Michie, van Stralen, & West, 2011) and the taxonomy of behavior change techniques (BCTs) (Michie et al., 2013). They found that self-monitoring, problem-solving, and restructuring the social or physical environment were particularly promising BCTs in reducing SB among adults.

In a recent systematic review (Chu et al., 2016), consistent evidence for intervention effectiveness was found for reducing the SB in workplace, particularly for multi-component (i.e., deploying sit-stand workstations in combination with behavioral interventions) and environmental strategies (i.e., using sit-stand workstation, portable elliptical/pedal machine, and stationary cycle ergometer and treadmill desk). Educational/behavioral strategies on their own (i.e., motivational interviewing, provide information on consequences of behavior to the individual, goal setting, action planning, prompt self-monitoring of behavior, provide instruction on how to perform the behavior, teach to use prompts/cues and facilitate social comparison), were less effective. The pooled intervention effect showed a significant workplace sitting reduction of \(-39.6\) min/8-h workday (95% confidence interval CI: \([-51.7, -27.5]\)), favoring the intervention group. Although this review found a significant result of effectiveness in SB reduction interventions, it did not compare the effectiveness of different behavior change techniques as shown in (Gardner et al., 2016), which is necessary for guiding intervention design.

### 2.2.2 Persuasive Technology

In the introduction, we can see the majority of interventions to reduce SB require changing environment or workplace policy, which are not available in many cases. We believe that the modern technology could provide promising tools for reducing SB in an effective and efficient way. Without a doubt, our lifestyle and daily behavior have been changed dramatically by modern technologies, such as computers and smartphones, and office workers spend most of their SB time using modern technologies and devices. Therefore, we seek ways of reducing prolonged SB from technology, which we call persuasive technology (PT).

The term, **persuasive technology**, describes technologies designed to change users’ attitude and behavior (Fogg, 1998). Fogg pointed out three methods technologies can change people: as tools, as media, and as social actors. Based on this understanding, Oinas-Kukkonen and Harjumaa (Oinas-kukkonen & Harjumaa, 2009) developed a framework called Persuasive System Design (PSD) that can be
2.2 Introduction

directly applied to persuasive system development. The PSD framework describes 28 persuasive technology principles in 4 categories (supporting primary task, computer-human dialogue, system credibility, and social) and explains how to transfer these principles into software functionality.

The PSD has been used explicitly and implicitly in health behavior change intervention studies (Lehto & Oinas-Kukkonen, 2015; Matthews, Win, Oinas-Kukkonen, & Freeman, 2016). Kelders et al. (Kelders, Kok, Ossebaard, & Van Gemert-Pijnen, 2012) provided a systematic review of the impact of the PSD on adherence to web-based interventions. Wildeboer and colleagues (Wildeboer, Kelders, & van Gemert-Pijnen, 2016) conducted a meta-analysis showing that web-based interventions with the principles in the PSD model have a large and significant effect size on mental health, and increasing the number of principles in different categories does not necessarily lead to better outcomes. In addition, they also found a number of combinations of principles that were more effective, e.g., tunneling and tailoring, reminders and similarity, social learning and comparison.

Unlike behavior change theories, e.g., Transtheoretical Model of Change (TTM) (Prochaska & Velicer, 1997) and the Health Belief Model (HBM) (Champion, 1984), which explain and predict behavior, persuasive technologies describe the characteristics how information systems should deliver behavior change interventions. Another confusion would be the relationship between persuasive technology and behavior change techniques (BCTs), which is a well-known taxonomy containing 93 items (e.g., goal-setting). As rooted from behavior change theories, BCTs provide the content for behavior change interventions. Although some elements (e.g., self-monitoring and rewards) appear in both BCTs and PSD principles, they are derived from different perspectives. We believe persuasive technologies should be used with proper behavior change theories in practice.

2.2.3 Aim

To our knowledge, there has been no systematic review of the impact of persuasive technology on interventions of reducing prolonged SB at work. Our aim, therefore, is to provide exactly this review of PT in the domain of SB change at work. In the rest of this chapter, we will first explain our systematic review process and the results. Then we show the analysis of the reviewed studies and the pitfalls for further studies. Finally, we give the conclusion and future work.
2.3 Methods

2.3.1 Data Source

Articles were searched using Google Scholar, ACM digital library, JMIRe, and PubMed5. The terms and the term combination strategy used to search target articles are listed in Table 2.1. All articles were published between 1987 and November 2016. The identified articles were further manually searched for relevant publications.

Table 2.1: The terms and combination strategy for searching articles.

<table>
<thead>
<tr>
<th>Number</th>
<th>Term</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Workplace</td>
<td>(1 OR 2 OR 3 OR 4) AND (5 OR 6)</td>
</tr>
<tr>
<td>2</td>
<td>Occupation</td>
<td>4) AND (5 OR 6)</td>
</tr>
<tr>
<td>3</td>
<td>At work</td>
<td>AND (7 OR 8 OR 9)</td>
</tr>
<tr>
<td>4</td>
<td>Office</td>
<td>9) AND 10</td>
</tr>
<tr>
<td>5</td>
<td>Sedentary</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Sitting</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Adults</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Worker</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Employee</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Intervention/s</td>
<td></td>
</tr>
</tbody>
</table>

2.3.2 Study Selection

Relevant papers were imported into Mendeley6 Desktop software, and duplicates were removed. The four-phase flow diagram of PRISMA (Liberati et al., 2009) was used to illustrate the study selection process. We filtered the interventions following the criteria:

- Target Group: Only adults who have sedentary lifestyle at work.
- Target Behavior: Only interventions aiming to reduce prolonged sitting behavior at work.
- Study Design: Only studies with a parallel control group.
- Measurement: Only studies reporting sitting time output objectively measured by activity tracker or self-report.
- Language: Only articles written in English.
- Persuasive Technology: Only interventions/intervention arms integrating PSD principles as variables.

---

4 http://www.jmir.org/
5 https://www.ncbi.nlm.nih.gov/pubmed
6 https://www.mendeley.com
2.3 Methods

2.3.3 Data Coding

All the selected interventions were coded according to the PSD principles (Oinas-kukkonen & Harjumaa, 2009), as shown in Error! Reference source not found..

Table 2.2: PSD principles.

<table>
<thead>
<tr>
<th>PSD principle</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Task Support</strong></td>
<td></td>
</tr>
<tr>
<td>Reduction (1.1)</td>
<td>System should reduce steps users take when performing target behavior.</td>
</tr>
<tr>
<td>Tunneling (1.2)</td>
<td>System should guide users in attitude/behavior change process by providing means for action.</td>
</tr>
<tr>
<td>Tailoring (1.3)</td>
<td>System should provide tailored info for user groups.</td>
</tr>
<tr>
<td>Personalization (1.4)</td>
<td>System should offer personalized content and services for individual users.</td>
</tr>
<tr>
<td>Self-monitoring (1.5)</td>
<td>System should provide means for users to track their performance or status.</td>
</tr>
<tr>
<td>Simulation (1.6)</td>
<td>System should provide means for observing link between cause &amp; effect with regard to users’ behavior.</td>
</tr>
<tr>
<td>Rehearsal (1.7)</td>
<td>System should provide means for rehearsing target behavior.</td>
</tr>
<tr>
<td><strong>Dialogue Support</strong></td>
<td></td>
</tr>
<tr>
<td>Praise (2.1)</td>
<td>System should use praise to provide user feedback based on behaviors.</td>
</tr>
<tr>
<td>Rewards (2.2)</td>
<td>System should provide virtual rewards for users to give credit for performing target behavior.</td>
</tr>
<tr>
<td>Reminders (2.3)</td>
<td>System should remind users of their target behavior while using the system.</td>
</tr>
<tr>
<td>Suggestion (2.4)</td>
<td>System should suggest users carry out behaviors while using the system.</td>
</tr>
<tr>
<td>Similarity (2.5)</td>
<td>System should imitate its users in some specific way.</td>
</tr>
<tr>
<td>Liking (2.6)</td>
<td>System should have a look &amp; feel that appeals to users.</td>
</tr>
<tr>
<td>Social role (2.7)</td>
<td>System should adopt a social role.</td>
</tr>
<tr>
<td><strong>System Credibility Support</strong></td>
<td></td>
</tr>
<tr>
<td>Trust-worthiness (3.1)</td>
<td>System should provide info that is truthful, fair &amp; unbiased.</td>
</tr>
<tr>
<td>Expertise (3.2)</td>
<td>System should provide info showing knowledge, experience &amp; competence.</td>
</tr>
</tbody>
</table>
Table 2.2: PSD principles.

<table>
<thead>
<tr>
<th>PSD principle</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface credibility (3.3)</td>
<td>System should have competent and truthful look &amp; feel.</td>
</tr>
<tr>
<td>Real-world feel (3.4)</td>
<td>System should provide info of the organization/actual people behind it content &amp; services.</td>
</tr>
<tr>
<td>Authority (3.5)</td>
<td>System should refer to people in the role of authority.</td>
</tr>
<tr>
<td>Third-party endorsements (3.6)</td>
<td>System should provide endorsements from external sources.</td>
</tr>
<tr>
<td>Verifiability (3.7)</td>
<td>System should provide means to verify accuracy of site content via outside sources.</td>
</tr>
<tr>
<td>Social Support</td>
<td></td>
</tr>
<tr>
<td>Social learning (4.1)</td>
<td>System should provide means to observe others performing their target behaviors.</td>
</tr>
<tr>
<td>Social comparison (4.2)</td>
<td>System should provide means for comparing performance with the performance of others.</td>
</tr>
<tr>
<td>Normative influence (4.3)</td>
<td>System should provide means for gathering people who have same goal &amp; make them feel norms.</td>
</tr>
<tr>
<td>Social facilitation (4.4)</td>
<td>System should provide means for discerning others who are performing the behavior.</td>
</tr>
<tr>
<td>Cooperation (4.5)</td>
<td>System should provide means for co-operation.</td>
</tr>
<tr>
<td>Competition (4.6)</td>
<td>System should provide means for competing with others.</td>
</tr>
<tr>
<td>Recognition (4.7)</td>
<td>System should provide public recognition for users who perform their target behavior.</td>
</tr>
</tbody>
</table>

2.4 Results

2.4.1 Search and Selection Results

A total of 1025 articles were identified from online database searching and other records, of which 38 articles were duplicated. Two coders screened the records separately and then resolved the conflicts together. Afterward, 708 records were screened out since no clear information relevant to the research topic was found. Finally, eight intervention studies were selected after full-text reading (see Figure 2.1). One of the coders coded the interventions and the other two coders checked the coding list. Differences were resolved by discussion.
2.4 Results

<table>
<thead>
<tr>
<th>Identification</th>
<th>Screening</th>
<th>Eligibility</th>
<th>Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records identified through database searching (n = 1018) Google Scholar = 924 ACM DL = 12 JMIR = 5 PubMed = 77</td>
<td>Records after duplicates removed (n = 987)</td>
<td>Records screened (n = 987)</td>
<td>Full-text articles assessed for eligibility (n = 279)</td>
</tr>
<tr>
<td>Additional records identified through other sources (n = 7)</td>
<td>Records excluded (n = 708) Title and abstract show no clear information relevant to this topic</td>
<td>Full-text articles excluded (n = 271) Not for office workers (n = 21) No intervention study (n = 93) No sitting outcome (n = 104) Duplicated Study (n = 9) Master/PhD Thesis (n = 14) No Persuasive Technology variables (n = 30)</td>
<td>Studies included in qualitative synthesis (n = 8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Studies included in quantitative synthesis (meta-analysis) (n = 8)</td>
</tr>
</tbody>
</table>

Figure 2.1: The study selection workflow.

2.4.2 Intervention Studies Description

All of the selected studies were conducted in developed countries within the last five years. Six out of eight studies are Randomized Controlled Trials (RCTs), while the other two are quasi-experiments. Only participants who finished the intervention duration are included. The sample size ranged from 28 to 246 and the interventions lasted from 5 days to 12 months. SB was objectively measured using activPAL or ActiGraph activity trackers in most of the studies (n = 6), and three studies used self-reporting questionnaire (IPAQ and WSQ) to estimate the physical activity and sedentary behavior. Details are listed in Table 2.3.

The aims of the interventions are all to reduce office workers' sedentary behavior during work and/or waking hours, but with different proximal aims, e.g., increasing standing time, short walks, or breaks. Four studies used no-treatment control conditions, while the other studies compared traditional intervention plus PT-intervention with a traditional intervention-only comparator arm. Three studies examined the effectiveness of hourly PC-based prompt to reduce sitting time at work (Donath, Faude, Schefer, & Roth, 2015; Evans, Fawole, Sheriff, & Dall, 2012; Wildeboer et al., 2016).
As mentioned in (Chu et al., 2016), organizational and environmental factors also impact sedentary behavior at work. Even though we focus on a specific measurement, such as total sitting time, the effect size may also have a large bias based on different intervention-control settings. Both (Donath et al., 2015) and (Evans et al., 2012) used a PC-based prompt to reduce sitting time at work. But in (Evans et al., 2012) all participants attended an education session, while in (Donath et al., 2015) participants in both groups were facilitated with height-adjustable working desks. Only three intervention studies contained a no-treatment comparator arm, which can only lead to less powerful meta-analysis.

2.4.3 Effects of Interventions

Four interventions (De Cocker et al., 2016; Evans et al., 2012; Gilson, Ng, Pavey, Ryde, & Straker, 2016; Puig-Ribera & Martínez-Lemos, 2015) showed statistically significantly positive effect on reducing sedentary behavior in at least one outcome (e.g., prolonged sitting time), compared with control groups. In another study (Brakenridge & Fjeldsoe, 2016) both the intervention group and control group showed significant improvement in SB. Among the remaining studies, one study (Donath et al., 2015) reported a notable but not significant decrease in sitting time and increase in standing time compared to the control condition, while the other one found a significant increase of weekday sedentary behavior with a small effect size in the intervention group. More details are listed in Table 2.4.

2.4.4 Persuasive Technology Analysis

Reminders (i.e., reminding participants of their goals to reduce the SB) was the most frequently used PSD principle among the reviewed studies, while tunneling (i.e., guiding participants to reduce SB by PC-based prompt, website, or email), self-monitoring (i.e., providing means to keep track of one’s own SB in real-time), and tailoring (i.e., providing tailored information to specific participant group or intervention phase) appeared in three studies (see Figure 2.2). Reflecting to the four function categories, primary task support (i.e., tunneling, tailoring, personalization, and self-monitoring) and dialogue support (i.e., reminders and suggestion) were often utilized, while social support and system credibility support appeared not as frequently, as only social comparison (i.e., sharing participants’ experience through social networks) and expertise (i.e., informing users about the health risks of prolonged SB) were used in these two categories.
2.4 Results

![Figure 2.2: PSD principles (see details in Error! Reference source not found.) usage frequency.](image)

2.4.5 In-Depth Analysis of Reminders

As six reviewed studies used reminders to inform participants of their sedentary behavior and timely breaking moment, it is of interest to make an in-depth analysis on it. We list the details of these reminders based on the reminder properties including frequency, interface, and content in Table 2.5.

2.4.5.1 Reminders Frequency

Two reviewed interventions (Taylor et al., 2016; Urda, Lynn, & Larouere, 2016) examined the effectiveness of hourly PC-based prompts on reducing occupational sitting along with tunneling or suggestion found no significant effect compared with the control groups. However, another intervention (Evans et al., 2012) adopted similarly frequent (every 30 min) PC reminders showed significant effect following a brief education session. Only one intervention (Gilson et al., 2016) used event-based reminders (i.e., every 30-60 min continuous sitting), which also conducted a participatory workshop before the intervention and resulted with a significant and larger effect on reducing SB than controlled group.

2.4.5.2 Prompts Interface

In most of the interventions (Donath et al., 2015; Evans et al., 2012; Gilson et al., 2016; Taylor et al., 2016; Urda et al., 2016), reminders were delivered via PC prompts with text messages, while only one via the smartphone (Brakenridge & Fjeldsoe, 2016). The effect of the interface cannot be evaluated based on the results of the reviewed studies because there are no enough controlled conditions to compare the different effect of this variable.
Table 2.3: Intervention study characteristics.

<table>
<thead>
<tr>
<th>Author, year, reference No.</th>
<th>Country</th>
<th>Study design</th>
<th>Sample size (finished)</th>
<th>Study duration</th>
<th>Measurement method</th>
<th>Intervention-control Description</th>
<th>Intervention Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evans et al. (2012)</td>
<td>UK</td>
<td>RCT</td>
<td>28 (I:14 C:14)</td>
<td>5 work days for baseline, 5 work days for intervention</td>
<td>objectively measured using the activPAL activity monitor</td>
<td>C: a brief education session on the importance of reducing long sitting periods at work I: same education along with prompting software on PC reminding to stand up every 30 minutes.</td>
<td>To reduce long uninterrupted sedentary periods and total sedentary time at work.</td>
</tr>
<tr>
<td>Donath et al. (2015)</td>
<td>Switzerland</td>
<td>RCT</td>
<td>31 (I: 15 C: 16)</td>
<td>5 work days for baseline, 12 weeks for intervention</td>
<td>objectively measured by ActiGraph wGT3X-BT</td>
<td>C: height-adjustable working desk I: height-adjustable working desk and three daily screen-based prompts</td>
<td>To increases daily office standing time in healthy middle-aged office workers.</td>
</tr>
<tr>
<td>Puig-Ribera et al. (2015)</td>
<td>Spain</td>
<td>Quasi-experimental</td>
<td>190 (I:88 C:102)</td>
<td>5 work days for baseline, 19 weeks for intervention, 2 weeks for follow-up</td>
<td>objectively measured by pedometers, self-reported using IPAQ</td>
<td>C: daily logging of their steps and sitting time I: daily logging and a web-based program with automatic email delivering in a three-phase intervention</td>
<td>To decrease occupational sitting time through increased incidental movement and short walks.</td>
</tr>
<tr>
<td>Brakenridge et al. (2016)</td>
<td>Australia</td>
<td>RCT</td>
<td>153 (I: 66 C: 87)</td>
<td>12 months</td>
<td>objectively measured using ActivPAL monitors</td>
<td>C: only organizational support I: organizational support plus LUMOback activity tracker</td>
<td>To reduce sitting in office workers.</td>
</tr>
<tr>
<td>Taylor et al. (2016)</td>
<td>USA</td>
<td>RCT</td>
<td>106 (I: 59 C: 47)</td>
<td>6 months</td>
<td>self-reported using IPAQ</td>
<td>C: no intervention I: PC-based prompt</td>
<td>To improve physical and behavioral health outcomes.</td>
</tr>
</tbody>
</table>
Table 2.3: Intervention study characteristics.

<table>
<thead>
<tr>
<th>Author, year, reference No.</th>
<th>Country</th>
<th>Study design</th>
<th>Sample size (finished)</th>
<th>Study duration</th>
<th>Measurement method</th>
<th>Intervention-control Description</th>
<th>Intervention Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Cocker et al. (2016)</td>
<td>Belgium</td>
<td>RCT</td>
<td>93 (I1: 35 I2: 35 C: 23)</td>
<td>3 months</td>
<td>self-reported using WSQ, objectively measured using activPAL activity monitor</td>
<td>C: no intervention I1: automatic Web-based, generic information and suggestions I2: automatic Web-based, computer-tailored intervention</td>
<td>To reduce and interrupt sitting at work.</td>
</tr>
<tr>
<td>Gilson et al. (2016)</td>
<td>Australia</td>
<td>Quasi-experimental</td>
<td>57 (I: 24 C: 33)</td>
<td>5 months</td>
<td>objectively measured using GENEActive activity monitor</td>
<td>C: participatory workshop only I: participatory workshop and a chair sensor/software package (Sitting Pad) with real-time prompts</td>
<td>To reduce occupational sedentary exposure and increase physical activity.</td>
</tr>
<tr>
<td>Urda et al. (2016)</td>
<td>Australia</td>
<td>RCT</td>
<td>44 (I: 22 C:22)</td>
<td>2 weeks</td>
<td>objectively measured using activPAL 3</td>
<td>C: no intervention I: maintained behaviors during control week, but received hourly alerts on their computer during work hours in the intervention week</td>
<td>To reduce sitting time, increase sit-to-stand transitions, and improve perceived wellness in women with sedentary jobs.</td>
</tr>
</tbody>
</table>
Table 2.4: The effects of the intervention studies.

<table>
<thead>
<tr>
<th>Author, year, reference No.</th>
<th>PSD principle</th>
<th>Outcomes</th>
<th>Intervention Group Results</th>
<th>Controlled Group Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evans et al. (2012)</td>
<td>2.3</td>
<td>Total sitting time (minutes/day [%]) Number of sitting events (events/day [events/hour]) Number of prolonged sitting events (events/day [events/hour]) Duration of prolonged sitting events (hours/day [%])</td>
<td>Total sitting time reduction is not significant, but the prolonged sitting time reduction is significant (-48(84) min/day or -12.2% (19.3%) with 95% CI, p&lt;0.05).</td>
<td>No significant difference</td>
</tr>
<tr>
<td>Donath et al. (2015)</td>
<td>1.2 2.3</td>
<td>Standing time (hours per week) Sitting time (hours per week)</td>
<td>Standing time 7.2 (4.8) to 9.7 (6.6) hours/week with p=0.09, sitting time 29.4 (6.5) to 27.8 (10.7) hours/week with p=0.63.</td>
<td>No significant difference</td>
</tr>
<tr>
<td>Puig-Ribera et al. (2015)</td>
<td>1.2 1.3 1.5 2.4 4.2</td>
<td>Step (counts/day) Sitting time (minutes/day)</td>
<td>Daily occupational sitting reduced significantly -32.2(9) min/day p=0.007, and step counts increased +924(245) P&lt;0.001.</td>
<td>No significant difference</td>
</tr>
<tr>
<td>Brakenridge et al. (2016)</td>
<td>1.5 2.3</td>
<td>Work hours and Overall hours: Sitting (min/10 h workday) Prolonged sitting (min/10 h workday) Time between sitting bouts (min) Standing (min/10 h workday) Stepping (min/10 h workday) Step count (number of steps/10 h workday)</td>
<td>Only in work hours, significant improvement in sitting (~35.5 (25.3) min, p = 0.006), prolonged sitting (~45.7 (38.3) min, p = 0.019), standing (~27.4 (19.7) min, p = 0.007), and stepping (~9.1 (8.9) min, p = 0.045).</td>
<td>During work hours: significant improvement in sitting (~40.5 (20.4) min, p &lt; 0.001), prolonged sitting (~41.3 (26.5)min, p = 0.002), standing (~39.2 (18.3) min, p &lt; 0.001), and Time between sitting bouts (~1.7 (1.4) min, p = 0.019)</td>
</tr>
</tbody>
</table>
Table 2.4: The effects of the intervention studies.

<table>
<thead>
<tr>
<th>Author, year, reference No.</th>
<th>PSD principle</th>
<th>Outcomes</th>
<th>Intervention Group Results</th>
<th>Controlled Group Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor et al. (2016)</td>
<td>1.2 2.3</td>
<td>Weekday Sedentary time (min/weekday) Weekend Sedentary time (min/weekend)</td>
<td>Weekday sedentary behavior has no significant change (P =.20), weekend sedentary behavior decreased significantly.</td>
<td>Weekday sedentary behavior increased significantly, weekend sedentary behavior has no significant change.</td>
</tr>
<tr>
<td>De Cocker et al. (2016)</td>
<td>1.3 1.4 1.5 2.4 3.2</td>
<td>Self-report Total sitting time (minutes/day) Self-report domain-specific sitting (minutes/day) objectively measured total sitting time awake (hours/day) objectively measured sitting at work (% work hours) objectively measured standing at work (% work hours) objectively measured breaks at work (No./work hours)</td>
<td>Significant decrease in self-reported total sitting (507 (104) to 425 (110) min/day, P&lt;0.001) and sitting at work (338 (107) to 259 (88) min/day, P&lt;0.001), and significant increase in objectively measured breaks at work (3.8(1.5) to 4.3(1.6) hours/day, P=0.09).</td>
<td>No significant difference</td>
</tr>
<tr>
<td>Gilson et al. (2016)</td>
<td>1.3 1.4 2.3</td>
<td>Sedentary behavior in % work time Light physical activity in % work time Moderate physical activity in % work time Total time sitting (min/day) Longest bout sitting (min/day)</td>
<td>Significant decrease in Sedentary (-8%, p = 0.012), increase in light physical activity (+8%, p=0.018), and decrease in longest bout sitting (-15 min).</td>
<td>Significant decrease in sedentary (-2%), increase in light physical activity (+1%) and increase the longest bout sitting (+17 min)</td>
</tr>
<tr>
<td>Urda et al. (2016)</td>
<td>2.3</td>
<td>during an 8.5-hour workday: Time sitting (hours/workday) sit-to-stand transitions</td>
<td>No significant difference in average sitting time and sit-to-stand transitions from baseline compared with intervention.</td>
<td>No significant difference in average sitting time and sit-to-stand transitions from baseline compared with intervention.</td>
</tr>
</tbody>
</table>
2.4.5.3 Reminders Contents

The contents of the reminders are not only the plain text to remind the time but also more complex sentences which reflect the other principles like suggestions (Donath et al., 2015; Taylor et al., 2016) and changing appearance like color (Gilson, Faulkner, Murphy, & Meyer, 2013). However, the contents seem to have no effect on the interventions because different contents lead to same results when comparing (Taylor et al., 2016) and (Urda et al., 2016). It is important to note that all of the contents in the reviewed studies are fixed, from which the users can get less information after several times appearance of the reminders.

<table>
<thead>
<tr>
<th>Study</th>
<th>Frequency</th>
<th>Interface</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evans et al. (2012)</td>
<td>Every 30 min.</td>
<td>PC prompt with text (enforced showing 1 min each time).</td>
<td>&quot;It's a break time.&quot;</td>
</tr>
<tr>
<td>Donath et al. (2015)</td>
<td>Three fixed times per day.</td>
<td>PC prompt with text (could be closed manually)</td>
<td>“Prolonged sitting is harmful!; Change your working position!; Lift up your working desk”</td>
</tr>
<tr>
<td>Taylor et al. (2016)</td>
<td>Hourly.</td>
<td>PC prompt</td>
<td>Tips to encourage users to get up and walk hallways, stairs, or outdoors.</td>
</tr>
<tr>
<td>Gilson et al. (2016)</td>
<td>Every 30-60 min continuous sitting.</td>
<td>PC prompt with a color indicator (from green to amber, and then to red).</td>
<td>Changing color.</td>
</tr>
<tr>
<td>Urda et al. (2016)</td>
<td>Hourly</td>
<td>PC prompt with audible alert.</td>
<td>&quot;Get up and move.&quot;</td>
</tr>
</tbody>
</table>

2.5 Pitfalls

2.5.1 PSD Principles Overlapping

When coding the intervention elements into PSD principles, we found that it is difficult to differ some similar meaning pairs, e.g., tailoring (1.3) and personalization (1.4), tunneling (1.2) and suggestion (2.4). The tailoring principle was defined as “Information provided by the system will be more persuasive if it is tailored to the potential needs, interests, personality, usage context, or other factors relevant to a user group,” while the personalization was “a system that offers personalized content or services has a greater capability for persuasion.”
So it is clear that the personalization should be a case of tailoring. Similarly, the tunneling principle and the suggestion principle are also overlapping. These overlaps can lead to confused coding. In this chapter, we regard the property allowing users to modify the interface as personalization and code multiple suggestions which can be followed step by step into tunneling.

2.5.2 Behavioral Theories

We regard sedentary behavior reduction as a health behavior change problem. However, among the selected intervention studies, only one (De Cocker et al., 2016) reported that it was based on behavior change theories (i.e., self-determination theory (Ryan & Deci, 2000), the theory of planned behavior (Ajzen, 1991), and self-regulation theory (Maes & Karoly, 2005)). Since evidence has shown that behavior change interventions based on theories of health behavior are more effective than the non-theory-based ones (R. Davis et al., 2015; Glanz & Bishop, 2010), more theory-integrated sedentary behavior reduction studies are expected. According to the related reviews (R. Davis et al., 2015; Glanz & Bishop, 2010), the most frequently used theories of behavior change include the Transtheoretical Model of Change (TTM) (Prochaska & Velicer, 1997), the Health Belief Model (HBM) (Champion, 1984), the Social Cognitive Theory (SCT) (Bandura, 1977a), and the Theory of Planned Behavior (TPB) (Ajzen, 1991), to name a few. We would encourage more research applying these theories or novel theories into this domain.

2.5.3 User Experience

No study measured user experience of the deployed persuasive technology after the interventions. The reason may be that most of the SB reduction studies were conducted by non-HCI researchers. However, it is essential to take the user experience into consideration when evaluating the persuasive technology. The user experience of short-term and long-term usage of the persuasive technology may reveal mediator effects between user acceptance and intervention effectiveness.

2.6 Implications

Based on our review results, we provide the following implications:

(1) Well-designed intervention studies (e.g., RCTs) on reducing prolonged sedentary behavior at work with explicit involvement of persuasive technology are still lacking. Therefore, we encourage interdisciplinary cooperation in this field.

(2) When applying PSD principles in intervention studies, the underlying behavior change theories are supposed to be reported explicitly. As social support in reviewed studies are rarely applied, it should be further explored in future studies.
(3) Researchers of persuasive technology or PSD should pay attention to the user experience of behavior change support systems that apply persuasive technologies.

(4) Based on the reviewed studies, only using PC-prompt reminders with tunneling or suggestions show no significance on reducing sedentary time, while combined with brief education session the reminders can significantly improve sedentary behavior. This can be a good option for corresponding public intervention designers and providers.

2.7 Conclusion

This systematic review is the first to examine the effectiveness of persuasive technology on reducing prolonged sedentary behavior. We also show that there is a lack of theory-based interventions and user experience considerations in the selected studies. However, given the small number of studies and inconsistent study design, we did not conduct a meta-analysis to analyze the correlation between Persuasive System Design principles and intervention outcomes.

Through the systematic review of the intervention studies to reduce sedentary behavior at work, we illustrated how the persuasive technology was used. We revealed that reminders are the most frequently used PSD principle, while the principles of system credibility support and social support were seldom deployed. An analysis based on the frequency, the interface and the contents of the reminders gave more insights on how was this most popular persuasive technology utilized. We also showed the pitfalls of the PSD principles and the reviewed interventions studies including the behavioral theories basis and user experience evaluation.

More intervention studies are expected to explicitly report the details following the PSD principles to make more powerful systematic review and meta-analysis. The research on the user experience of the persuasive technology delivered to reduce prolonged sedentary behavior at work should draw more attention from the HCI community. More theory-based behavior change interventions utilizing persuasive technology are required to enable comprehensive meta-analysis. More longitudinal studies are also required to evaluate the long-term effects of SB reduction interventions.

The future work would be more focused on the properties of the variety of persuasive technologies to design better interventions on reducing sedentary behavior.
3 INTEGRATING TAXONOMIES INTO THEORY-BASED DIGITAL HEALTH INTERVENTIONS FOR BEHAVIOR CHANGE

“无以规矩，不能成方圆。”

Nothing can be accomplished without norms or standards.

- 孟子 (372 BC – 289 BC)
  Mencius (A Chinese Philosopher)

---


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, proposed the framework, and spearheaded the writing. Ahmed Fadhil and Jan-Philipp Lange helped in discussing the framework, writing parts of the paper, and reviewing the paper. Harald Reiterer supervised the work and reviewed this paper.
3.1 Abstract

Digital health interventions have been emerging in the last decade. Due to their interdisciplinary nature, digital health interventions are guided and influenced by theories (e.g., behavioral theories, behavior change technologies, persuasive technology) from different research communities. However, digital health interventions are always coded using various taxonomies and reported in insufficient perspectives. The inconsistency and incomprehensiveness will bring difficulty for conducting systematic reviews and sharing contributions among communities. Based on existing related work, therefore, we propose a holistic framework that embeds behavioral theories, behavior change technique (BCT) taxonomy, and persuasive system design (PSD) principles. Including four development steps, two toolboxes, and one workflow, our framework aims to guide digital health intervention developers to design, evaluate, and report their work in a formative and comprehensive way. We use the proposed framework as guidelines in our intervention studies for sedentary behavior change (see Chapter 6 and Chapter 7) and the systematic review of technologies for health behavior change (see Chapter 9).

3.2 Introduction

With the potential for low cost and high scalability for chronic disease prevention, in the past decade, digital health interventions (DHIs) have been widely discussed by government stakeholders, clinicians, and researchers (WHO, 2018a). Designing and deploying DHIs are challenging due to the complexity of human behavior, which could be affected by individuals’ motivation, emotion, ability, social environment, and physical environment. Therefore, DHI design could accordingly require theories and practice from several disciplines, including phycology, public health, behavioral science, human-computer interaction, and so on. The interdisciplinary nature of DHIs calls for a comprehensive framework to guide the development, evaluation, and report.

As DHIs are expected to change human behavior, behavioral theories can serve as the development foundation. It has shown that theory-based behavior change interventions are more effective than others (R. Davis et al., 2015; Glanz & Bishop, 2010). Nevertheless, behavioral theories could also be ignored (as discussed in Chapter 2 (Y. Wang, Wu, et al., 2018)) or misused (Hekler et al., 2013). Although behavioral theories allow to explain and predict behavior, they lack the guidance of translating into operational techniques.

The Behavior Change Technique (BCT) taxonomy (Abraham & Michie, 2008) and persuasive technology design (PSD) principles (Oinas-kukkonen & Harjumaa, 2009) are two widely used taxonomy in DHIs research (Geuens et al., 2016; Kelders et al., 2012; Middelweerd, Mollee, van der Wal, Brug, & te Velde, 2014; Olander et al., 2013). These taxonomies not only inform DHI design but also
enable precise reporting, which will be favored by systematic reviewers. Although derived from different philosophies, BCTs and PSDs have some common elements. However, they are used separately in many DHI studies. To benefit from both, we combine the BCT taxonomy and PSD principles into a more comprehensive taxonomy in the light of the Behavioral Intervention Technology (BIT) model (Mohr, Schueller, Montague, Burns, & Rashidi, 2014).

In this chapter, we aim to put the puzzles together and build a holistic framework to aid DHIs researchers to design, evaluate, and report their studies. In short, our contributions include:

1. We provide a holistic framework that allows DHI developers to design, evaluate, and report their work in a formative and comprehensive way.
2. We classified PSD principles into two parts: strategies and characteristics. We then combine the BCT taxonomy and PSD principles (the characteristics) into our DHI taxonomy.

By elaborating the BIT model, we propose a comprehensive way to report DHI description: strategies, characteristics, and a workflow.

3.3 Related Work

As this chapter is for DHI developers from different communities, it is necessary to clarify the terms and our scope before we present the related work. Digital health or eHealth is the umbrella concept referring to the use of information and communication technologies (ICT) for health (WHO, 2018b). According to the world health organization (WHO), digital health interventions (DHIs) covers systematic functionalities to support clients, healthcare providers, health system or resource managers, and data services (WHO, 2018a). In this chapter, however, we limit our scope to the DHIs aiming to change users’ lifestyle behavior (e.g., food intake, physical activity, and smoking) using digital technology to prevent or manage health problems.

3.3.1 CeHRes Roadmap

Back in 2011, a holistic framework (i.e., CeHRes Roadmap) was proposed to improve the uptake and impact of eHealth technology. The CeHRes roadmap was built upon 16 existing frameworks via a systematic review and emphasized the importance of holism (van Gemert-Pijnen et al., 2011). Human characteristics, socioeconomic and cultural environments, and technology are closely connected to affect human behavior. Therefore, developers should always keep these holistic factors in mind to build eHealth technologies. Within this framework, CeHRes roadmap was illustrated as a practical guideline to help plan, coordinate, and execute the participatory development process of eHealth technologies. CeHRes roadmap consists of six steps - contextual inquiry, value specification, design, operationalization, and summative evaluation - which integrate persuasive technology design, human-centered design, and business...
3.3 Related Work

modeling. Although CeHRes roadmap integrates behavioral theories as for the foundation, it does not explicitly show how to apply them in the intervention design. Besides, CeHRes roadmap does not adopt any persuasive technology taxonomy.

3.3.2 Behavioral Intervention Technology Model

In 2014, Mohr and colleagues proposed the behavioral intervention technology (BIT) model aiming to support the translation of treatment and intervention-aims into an implementable treatment model (Mohr et al., 2014). The BIT model includes a theoretical phase followed by an instantiation phase. The theoretical phase consists of the intervention aims and behavior change strategies, whereas the instantiation level consists of intervention elements, characteristics, and workflow. Thus the BIT model can serve as a supplement to the CeHRes roadmap. However, the BIT model only provided some examples in each component. E.g., behavior change strategies include education, goal setting, monitoring, feedback, and motivation enhancement. As Mohr et al. mentioned, the BIT model is a simplification and should be modified and elaborated to fit users’ need (Mohr et al., 2014). In this chapter, we will adjust and elaborate the BIT model to fit into our holistic framework.

3.3.3 IDEAS

In 2016, Mummah et al. proposed IDEAS (Integrate, Design, Assess, and Share) as a framework and toolkit of strategies for the development of DHIs (Sarah Ann Mummah, Robinson, et al., 2016). IDEAS was built on three essential components: behavioral theory, design thinking, and evaluation and dissemination. The IDEAS framework emphasizes the importance of behavioral theories and introduces the taxonomy of behavior change techniques (BCTs). However, the BCT taxonomy is regarded as an alternative to using behavioral theories to identify target constructs in interventions. In our holistic framework, we suggest using both of them as two necessary steps because they correspond to the intervention aims and strategies respectively.

3.3.4 The Theory & Techniques Tool

Although the BCT taxonomy provides the active components for the design of behavior change interventions, understanding their links to the mechanisms of action (i.e., the constructs in behavioral theories (Michie et al., 2016)) could help inform the intervention development. To bridge this gap, Michie et al. developed the Theory and Techniques Tool. Following the initiative of Michie et al. (Michie et al., 2016) as well as integrating Fogg’s Behavior Model (Fogg, 2009a, 2009b), Wahl suggested an integrated framework (Wahl, 2019). However, Wahl’s framework aimed to guide the “in-the-moment” assessment of eating behavior change, instead of providing a comprehensive one for the design and report of health behavior change interventions. In this chapter, we integrate the strategies
(including BCTs), their characteristics (e.g., the media to deliver interventions), the workflow (including in-the-moment or “event-based”), and the underlining behavioral theories into a holistic framework.

### 3.4 Behavioral Theories

All the three reviewed work above mention behavioral theories, but only IDEAS explicitly integrate behavioral theories into the step development process. Behavioral theories refer to the social-psychological theories of behavior change, which explain and predict human behavior. As depicted by Sutton (Sutton, 2002), each of the behavioral theories specifies a small number of cognitive and affective factors as the proximal determinants of behavior (see Figure 3.1). These factors are called constructs in behavioral science (Hekler et al., 2013). We will use this term to refer to the fundamental components of behavioral theories in the rest of the chapter.

![Figure 3.1: A hypothesized continuum model. The constructs in black are borrowed from the integrated behavioral model in (Conner & Norman, 2005). The construct “planning” is from the Health Action Process Approach (HAPA) (Schwarzer, 2016), and the construct “habit” is added inspired by the work (Maher & Conroy, 2015).](image)

Glanz et al. (Glanz K, Rimer BK, 2008) illustrated the most frequently used behavioral theories published before 2010: the Social Cognitive Theory (SCT) (Bandura, 1977a), the Transtheoretical Model of Change (TTM) (Prochaska & Velicer, 1997), the Health Belief Model (HBM) (Rosenstock, Strecher, & Becker, 1988), and the Theory of Planned Behavior (TPB) (Ajzen, 1991). Davis et al. (Rachel Davis, Rona Campbell, Zoe Hildon, 2014) also identified 82 behavioral theories, among which the most frequently used theories are TTM, TPB, SCT, the Information-Motivation-Behavioral-Skills Model, HBM, the Self-determination Theory (Ryan & Deci, 2000), the Health Action Process Approach (HAPA)
(Schwarzer, 2008), and the Social Learning Theory (Bandura, 1977b). Based on different assumptions of human behavior, these behavioral theories can be grouped into continuum theories and stage theories (Conner & Norman, 2005).

Continuum theories assume people’s behavior is caused by a set of variables, e.g., intention and skills. Except for TTM, all other mentioned theories fall into this group. Based on the behavioral model integrating several existing ones (Conner & Norman, 2005), we present a hypothesized continuum model as shown in Figure 3.1. Planning (shown in red in Figure 3.1) is specified as a mediator of the intention-behavior relationship in HAPA (Schwarzer, 2008; Sutton, 2008). Habit (shown in green in Figure 3.1) has been found being able to moderate the effects of planning (Maher & Conroy, 2015).

Stage theories assume people change their behavior in a process including several stages. The factors pushing people from one stage to the next are believed to be different. Therefore, the strategies at each state should be adapted accordingly. E.g., Figure 3.2 shows the stages and strategies of TTM.

Stage theories assume people change their behavior in a process including several stages. The factors pushing people from one stage to the next are believed to be different. Therefore, the strategies at each state should be adapted accordingly. E.g., Figure 3.2 shows the stages and strategies of TTM.

![Figure 3.2: The Transtheoretical Model of Change (TTM), adapted from (American Society on Aging and American Society of Consultant Pharmacists Foundation; 2012). This model divides the behavior change process into five stages, namely precontemplation, contemplation, preparation, action, and maintenance and relapse prevention. Depending on the stage of change, different strategies could be applied accordingly to make the intervention effective.](image-url)

Behavioral theories provide a toolbox to understand human behavior and explain the rationale behind interventions. However, their shortcomings should
be noticed before they are used. Hekler and colleagues (Hekler et al., 2013) have pointed out three shortcomings of behavioral theories: (1) most behavioral theories explain only a small portion of variance in the outcomes they are trying to account for; (2) many behavioral theories, in their current form, are not falsifiable; and (3) there is a fragmentation and an over-abundance of different theories. Therefore, DHI developers should base on without being limited to behavioral theories. With the emerging of DHIs, the existing behavioral theories can be further evaluated and improved (Klasnja, Hekler, Korinek, Harlow, & Mishra, 2017). Here we list some guidelines when using specific behavioral theories: (Conner & Norman, 2005) and (Maberry, 2016) for the SCT, (Champion, 1984) for the HBM, (Ajzen, 2006) for the TBP, and (Schwarzer, 2019) for the HAPA.

3.5 Digital Health Intervention Taxonomy

While behavioral theories can predict and explain human behavior, there is a gap between theories and operational interventions. Will self-monitoring increases self-efficacy for promoting physical activity? Will information about health consequences affects perceived advantages/disadvantages? Due to the high complexity of human behavior and health, one DHI may involve several techniques. The lack of a consistent taxonomy of DHIs will lead to poor replicability and low comparability of the results from related studies. Although there exist taxonomies to bridge the theory-intervention gap, the use of different taxonomies still hinders the understanding and contribution among communities. Therefore, we present the DHI taxonomy, a unified taxonomy to take advantage of two widely used taxonomies (the BCT taxonomy and PSD principles) in light of the BIT model.

Behavior change techniques (BCTs) are defined as observable, replicable, and irreducible components of an intervention designed to change behavior (Abraham & Michie, 2008; Michie et al., 2013), e.g., self-monitoring or goal setting. Abraham and Michie developed a taxonomy of behavior change techniques, which identified 22 BCTs and 4 BCT packages (Abraham & Michie, 2008) and was later extended to a taxonomy containing 93 BCTs into 16 groups, called Behavior Change Technique Taxonomy (v1) (Michie et al., 2013). The BCT taxonomy has been used for informing intervention development and report (Sarah A. Mummah, King, Gardner, & Sutton, 2016; Sarah Ann Mummah, Mathur, King, Gardner, & Sutton, 2016) and identifying the effectiveness of BCTs (Dombrowski et al., 2012; Gardner et al., 2016; Michie, Abraham, Whittington, McAteer, & Gupta, 2009; Olander et al., 2013). It also provides a means to evaluate health and fitness apps (Conroy, Yang, & Maher, 2014; Crane, Garnett, Brown, West, & Michie, 2015; Direito et al., 2014; Middelweerd et al., 2014) and wearables (Lyons, Lewis, Mayrsohn, & Rowland, 2014). From the official website of the BCT taxonomy (BCT Taxonomy Research Team, 2014), we found a collection of 405 intervention studies with BCTs coding. We show the word cloud
of BCTs based on this collection in Figure 3.3. The top-five used or tested BCTs are goal setting (behavior), instruction on how to perform a behavior, problem-solving, information about health consequences, and action planning.

Figure 3.3: The word cloud of BCTs used in the study collection of 405 intervention studies from the official website of the BCT taxonomy (BCT Taxonomy Research Team, 2014).

In related work, we have introduced the behavioral intervention technology (BIT) model (Mohr et al., 2014). In terms of the intervention strategies in the BIT model, only some examples (i.e., education, goal setting, monitoring, feedback, and motivation enhancement) were provided. We think the BCT taxonomy can serve as a strategy pool for the BIT model.

Aiming to create a conceptual framework that can be directly applied to persuasive system development, the persuasive system design (PSD) model describes 28 principles in four categories (supporting primary task, computer-human dialogue, system credibility, and social) as an extension of Fogg’s work on persuasive technology (Fogg, 2003). Table 3.1 describes the details of PSD principles. We found there are 16 principles (highlighted in red in Table 3.1) that have the same or similar definitions with counterparts from the BCT taxonomy. For example, self-monitoring appears both in PSD principles and the BCT taxonomy. Tunneling (1.2) in PSD principles has the same meaning with the BCT “4.1 structure on how to perform the behavior”. Please refer to Multimedia Appendix 1 for more details. We could not find their counterparts from the BCT taxonomy of five PSD principles (highlighted in blue in Table 3.1), which can serve as a supplement of the BCT taxonomy.
**Table 3.1: PSD principles.** The red principles have counterparts with the same or similar definitions in the BCT taxonomy. The blue principles have no counterparts in the BCT taxonomy but can also be regarded as intervention strategies. The green principles are characteristics of intention media.

<table>
<thead>
<tr>
<th>PSD principle</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Task Support</strong></td>
<td></td>
</tr>
<tr>
<td>Reduction (1.1)</td>
<td>System should reduce steps users take when performing target behavior.</td>
</tr>
<tr>
<td>Tunneling (1.2)</td>
<td>System should guide users in attitude/behavior change process by providing means for action.</td>
</tr>
<tr>
<td>Tailoring (1.3)</td>
<td>System should provide tailored info for user groups.</td>
</tr>
<tr>
<td>Personalization (1.4)</td>
<td>System should offer personalized content and services for individual users.</td>
</tr>
<tr>
<td>Self-monitoring (1.5)</td>
<td>System should provide means for users to track their performance or status.</td>
</tr>
<tr>
<td>Simulation (1.6)</td>
<td>System should provide means for observing link between cause &amp; effect with regard to users’ behavior.</td>
</tr>
<tr>
<td>Rehearsal (1.7)</td>
<td>System should provide means for rehearsing target behavior.</td>
</tr>
<tr>
<td><strong>Dialogue Support</strong></td>
<td></td>
</tr>
<tr>
<td>Praise (2.1)</td>
<td>System should use praise to provide user feedback based on behaviors.</td>
</tr>
<tr>
<td>Rewards (2.2)</td>
<td>System should provide virtual rewards for users to give credit for performing target behavior.</td>
</tr>
<tr>
<td>Reminders (2.3)</td>
<td>System should remind users of their target behavior while using the system.</td>
</tr>
<tr>
<td>Suggestion (2.4)</td>
<td>System should suggest users carry out behaviors while using the system.</td>
</tr>
<tr>
<td>Similarity (2.5)</td>
<td>System should imitate its users in some specific way.</td>
</tr>
<tr>
<td>Liking (2.6)</td>
<td>System should have a look &amp; feel that appeals to users.</td>
</tr>
<tr>
<td>Social role (2.7)</td>
<td>System should adopt a social role.</td>
</tr>
<tr>
<td><strong>System Credibility Support</strong></td>
<td></td>
</tr>
<tr>
<td>Trust-worthiness (3.1)</td>
<td>System should provide info that is truthful, fair &amp; unbiased.</td>
</tr>
<tr>
<td>Expertise (3.2)</td>
<td>System should provide info showing knowledge, experience &amp; competence.</td>
</tr>
<tr>
<td>Surface credibility (3.3)</td>
<td>System should have competent and truthful look &amp; feel.</td>
</tr>
<tr>
<td>Real-world feel (3.4)</td>
<td>System should provide info of the organization/actual people behind it content &amp; services.</td>
</tr>
<tr>
<td>Authority (3.5)</td>
<td>System should refer to people in the role of authority.</td>
</tr>
<tr>
<td>Third-party endorsements (3.6)</td>
<td>System should provide endorsements from external sources.</td>
</tr>
<tr>
<td>Verifiability (3.7)</td>
<td>System should provide means to verify accuracy of site content via outside sources.</td>
</tr>
<tr>
<td><strong>Social Support</strong></td>
<td></td>
</tr>
<tr>
<td>Social learning (4.1)</td>
<td>System should provide means to observe others performing their target behaviors.</td>
</tr>
<tr>
<td>Social comparison (4.2)</td>
<td>System should provide means for comparing performance with the performance of others.</td>
</tr>
<tr>
<td>Normative influence (4.3)</td>
<td>System should provide means for gathering people who have same goal &amp; make them feel norms.</td>
</tr>
<tr>
<td>Social facilitation (4.4)</td>
<td>System should provide means for discerning others who are performing the behavior.</td>
</tr>
<tr>
<td>Cooperation (4.5)</td>
<td>System should provide means for cooperation.</td>
</tr>
<tr>
<td>Competition (4.6)</td>
<td>System should provide means for competing with others.</td>
</tr>
<tr>
<td>Recognition (4.7)</td>
<td>System should provide public recognition for users who perform their target behavior.</td>
</tr>
</tbody>
</table>
Next, we present the diagram of our DHI taxonomy (see Figure 3.4). We just have shown its strategy part, which includes 93 (from the BCT taxonomy) plus 5 (real-world feel, verifiability, cooperation, competition, and recognition from PSD principles) strategies. The other part of our DHI taxonomy corresponds to the characteristics. The BIT model described four characteristics (medium, complexity, aesthetics, and personalization). Inspired by the characteristics related PSD principles (highlighted in green in Table 3.1), we include social role and trustiness, in addition to the mentioned four from the BIT model, into the characteristics part of the DHI taxonomy.

We divided the PSD principles into two groups. The ones fitting the definition of the BCT go to the strategies group, while others fall into characteristics group. Personalization is one of the characteristics in the BIT model. We find that tailoring has very close meaning to personalization according to their definitions in the PSD principles (Oinas-Kukkonen, 2013). We argue that similarity is also in line with the definition of personality. Therefore, we regard both tailoring and similarity the same as personality. Likewise, trust-worthiness and surface credibility are merged to one characteristic as trustiness. By dividing the PSD principles and merging the overlapping principles, we hope our new taxonomy can reduce the confusion and difficulty of coding DHIs (Geuens et al., 2016; Y. Wang, Wu, et al., 2018). Please see Appendix 2 for the complete list of elements in our DHI taxonomy.

![Figure 3.4: The diagram of our DHI taxonomy. The blue part is the strategy part, while the green part is the characteristics part.](image)

3.6 The Holistic Framework

The proposed holistic framework (see Figure 3.5) is called TUDER (Targeting, Understanding, Designing, Evaluating and Refining), which consists of four steps, two toolboxes (behavioral theories and the DHIs taxonomy), and a workflow. In each step, it is allowed to go back and update corresponding information.
Targeting the user group, the health problem, and the behavior. The target group, health problem, and behavior define the intervention aim(s). For example, an intervention to promote the use of standing desks (the behavior) to reduce the prolonged sedentary behavior (the behavior) of office workers (the user group) to prevent chronic diseases, e.g., type 2 diabetes (the health problem) (De Cocker, Veldeman, et al., 2015). The intervention designers should explain the relationship between the health problem and the behavior. Scientific evidence provides the rationale. E.g., the evidence that sedentary behavior and moderate-to-vigorous physical activity are independently associated with clustered cardiometabolic health supports the development of interventions to reduce office workers’ sedentary behavior (Knaeps et al., 2016). Another example is about myopia among children. A study showed that the time of outdoor activities was the most significant factor of myopia in 6- and 7-year-old Chinese children (Rose et al., 2008). Therefore, a reasonable intervention to reduce myopia (the health problem) among children (the user group) would be increasing their outdoor activity time (the behavior). Besides the scientific support, another rule is about the measurability to enable quantitative analysis. The target health problem is not necessarily measurable in an intervention study, while the target behavior must be (Wendel, 2013).

Understanding the mechanism underlining human behavior. Behavioral theories (e.g., see Figure 3.1 and Figure 3.2) provide DHI developers a toolbox to understand human behavior. Given the target user group, health problem, and behavior, developers ought to take one behavioral theory or a set of constructs from behavioral theories as the base of intervention design in the following step. We suggest that theory-based interventions should relate their strategies to specific constructs from behavioral theories. For example, an intervention design based on HAPA intended to support action planning (the construct) to reduce users’ sedentary behavior (Y. Wang, Fadhil, & Reiterer, 2018). Therefore, in addition to measuring the sedentary behavior, the constructs in HAPA should also be
3.6 The Holistic Framework

assessed. When analyzing the invention effect on action planning, the assessment of action planning is enough. However, in the case of analyzing the intervention effect on sedentary behavior, other constructs besides action planning have also to be considered. The participants should be grouped based on the level of their intention in data analysis. Alternatively, the user group in the previous step can be adjusted to only focus on one user group with a specific level of intention. During this step, DHI developers may backtrack to the previous step to adjust the target user group and measurements.

Designing the intervention strategies, characteristics, and workflow. We have included 98 intervention strategies and six characteristics in our DHI taxonomy. DHI developers can select a set of strategies based on their idea and describe the characteristics of their strategies according to the DHI taxonomy. As the context of an intervention may vary over time, the workflow that allows an intervention to be delivered according to time, task, or event would be demanding (Mohr et al., 2014). The workflow design has been comprehensively illustrated in the BIT model (Mohr et al., 2014) and the Just-in-Time Adaptive Intervention (JITAI) framework (I Nahum-Shani, Smith, & Tewari, 2014). From the perspective of implementation difficulty, time-based workflow (e.g., an hourly reminder in sedentary behavior intervention as discussed in Chapter 2 (Y. Wang, Wu, et al., 2018)) is the easiest. Task-based (e.g., a set of interventions delivered to a user sequentially) or event-based (e.g., adaptive food recommendation according to a user’s previous meal) workflow requires user data input. Because of the difficulty of inquiring users’ context data, the research on opportune moments for DHIs is still in the early stage (Sano, Johns, & Czerwinski, 2017; Y. Wang, Duan, Mueller, Butscher, & Reiterer, 2016).

Evaluating and refining the intervention design. Intervention evaluation could include usability evolution (regarding human-computer interaction) an effectiveness evaluation (regarding behavior change) in correspond to the uptake and impact of the intervention respectively (van Gemert-Pijnen et al., 2011). Think-aloud (Joe, Chaudhuri, Le, Thompson, & Demiris, 2015) and cognitive walkthrough can be used in the early stage of ideation creation and prototype to identify the usability issues. Then a pilot study with a small number of participants would be deployed to test the feasibility of the whole study procedure. Because many interventions need field study, the pilot can also help find some unknown issues in real-world use. Finally, heuristic evaluations based on randomized controlled trials (RCTs) (Sarah Ann Mummah, Robinson, et al., 2016) or sequential multiple assignment randomized trials (SMARTs) (Collins, Murphy, & Strecher, 2007) have to be conducted to generate powerful results. In our framework, an iterative evaluation and refinement process is adopted. Because evaluation and refinement are always intertwined with each other, we place them in one step in our framework.
3.7 Discussion

We have described the TUDER, a holistic framework to guide digital health intervention (DHI) development. We also provide a checklist for DHI developers, as shown in Figure 3.6. By completing the checklist and reporting all the details of a DHI study, the data coding work in systematic reviews could be reduced much.

<table>
<thead>
<tr>
<th>Targeting</th>
<th>Understanding</th>
<th>Designing</th>
<th>Evaluating and Refining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target user group:_____</td>
<td>Behavioral theories:__</td>
<td>Strategies:________</td>
<td>Study design:________</td>
</tr>
<tr>
<td>Target disease:_______</td>
<td>Constructs:________</td>
<td>Characteristics:_____</td>
<td>Evaluation results:_____</td>
</tr>
<tr>
<td>Target behavior:_______</td>
<td>Other factors:_______</td>
<td>Workflow:________</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.6: The checklist for using TUDER.

We build TUDER based on several existing related work (e.g., (Abraham & Michie, 2008; Mohr et al., 2014; Sarah Ann Mummah, Robinson, et al., 2016; Oinas-kukkonen & Harjumaa, 2009; van Gemert-Pijnen et al., 2011)). The key contribution of this work is to embed behavioral theories, behavior change technology (BCT) taxonomy, and persuasive system design (PSD) principles into a holistic framework. We believe this framework will be beneficial to each of them. This holistic framework and the DHI taxonomy will also enable more research questions. We provide some examples as follows:

1. What combinations of DHI strategies, characteristics, and workflow work better than others? In (Wildeboer et al., 2016), a meta-analysis shows several combinations of PSD principles were more effective, e.g., tunneling and tailoring, reminders and similarity, social learning and comparison. With consideration of the characteristics and workflow when coding the interventions, the results of intervention effectiveness analysis may change.

2. Is the DHI taxonomy able to explain more variance in DHI adherence? Kelders et al. (Kelders et al., 2012) systematically reviewed the impact of the PSD principles on adherence to web-based interventions. Their model explained 55% of the variance in users’ adherence. The DHI taxonomy brings more perspectives to analyze the effects of the components in interventions.

As the TUDER framework is expected to be comprehensive, some parts are simplistic. For example, only several behavioral theories are discussed. The DHI taxonomy is built upon two existing taxonomies. The DHI developers who are not familiar with the BCT taxonomy and PSD principle will find it challenging to use the DHI taxonomy.

3.8 Conclusion

This work presented the TUDER framework, containing four steps (targeting, understanding, designing, evaluating and refining), two toolboxes (behavioral
3.8 Conclusion

theories and digital health intervention taxonomy), and a workflow. The framework aims to integrate the advantages of behavioral theories, behavior change technique taxonomy, and persuasive technology design principles. Thus, it can help the digital health intervention researchers to design, evaluate, and report their studies in a formative and comprehensive way. By using this framework, future systematic reviews could have broader insights into digital health intervention studies. To better bridge the research from different communities, we will continue to test and improve this framework.
4 UNDERSTANDING USERS’ MOBILITY PATTERNS: CLUSTERING

横看成岭侧成峰。

From a point of view, it’s a hill; from another, it’s a mountain.

- 苏轼 (1037 – 1101)
  SU Shi
  (A Chinese writer, poet, painter, calligrapher, pharmacologist, gastronome, and a statesman of the Song dynasty)


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Björn Sommer helped in discussing the study design and data analysis, and reviewing the paper. Falk Schreiber and Harald Reiterer supervised the work and reviewed this paper.
4.1 Abstract

In this chapter we propose a method for extracting significant places or places of interest (POIs) using individuals’ spatio-temporal data for human mobility analysis, which we used for extracting sedentary location in our intervention study (see Chapter 6). General clustering methods such as DBSCAN were often used for detecting POIs from human mobility data. However, these methods do not use the temporal information in human mobility data well, and the only applied temporal information in related work is the time interval of consecutive location data for stay-point detection. Considering temporal constraints in human mobility, we propose a POI clustering approach – namely POI clustering with temporal constraints (PC-TC) – to extract POIs from spatio-temporal data of human mobility. Following the nature of human mobility in modern society, our approach aims to extract both global POIs (e.g., workplace or university) and local POIs (e.g., library, lab, and canteen). Based on two publicly available datasets including 193 individuals, our evaluation shows that PC-TC has advanced features in POI granularity and the potential of sequential POI prediction. We also tested PC-TC in a real-world mobile application: the user study results show high precision of our method for POI extraction in the university environment.

4.2 Introduction

Human mobility patterns can reflect lifestyles and routines, which influence individuals’ health. Chronic conditions - such as cardiovascular diseases and diabetes - place a heavy burden on individuals and the whole society. Unhealthy lifestyle choices - like smoking, physical inactivity, and unbalanced food intake - are highly related to these chronic diseases. Accordingly, an increasing number of studies have been focusing on personalized health behavior change to help individuals to prevent the onset of chronic diseases (Ma, Rosas, & Lv, 2016; Payne, Lister, West, & Bernhardt, 2015). For instance, the just-in-time adaptive interventions framework emphasizes context detection to provide personalized interventions for behavior change (Inbal Nahum-Shani et al., 2016). In recent studies, location-based interventions have shown great potential in mobile health applications (Naughton et al., 2016; Rabbi, Aung, Zhang, & Choudhury, 2015). For example, MyBehavior (Rabbi et al., 2015) generates personalized activity suggestions tailored to different places according to users’ activity levels. To improve location-based applications, in this chapter, we focus on extracting places of interest (POIs, see Sect. 3.2 for a formal definition) of human mobility.

POI extraction in human mobility analysis usually applies spatio-temporal data clustering as the key technique (Ikanovic & Mollgaard, 2017; Montoliu, Blom, & Gatica-Perez, 2013; Scellato, Musolesi, Mascolo, Latora, & Campbell, 2011; Sheng et al., 2016; Ye, Zheng, Chen, Feng, & Xie, 2009). Classical clustering methods include hierarchical clustering (e.g., Linkage (Sibson, 1972)), partitional
clustering (e.g., k-means (Macqueen, 1967)), grid-based clustering (Zheng, Zheng, Xie, & Yang, 2010), and density-based clustering (e.g., DBSCAN (Ester, Kriegel, Sander, & Xu, 1996)). One challenge of applying these clustering methods is to determine suitable parameters, which are usually chosen heuristically (e.g., 50 meters as the minimum distance between POIs) or by using optimization techniques (e.g., Silhouette Coefficient (Rousseeuw, 1987), Davis-Bouldin index (Davies & Bouldin, 1979), and the reachability-based method (Sander, Qin, Lu, Niu, & Kovarsky, 2003)). As an optimization problem, the clustering process is to obtain high intra-cluster similarity and low inter-cluster similarity (Rokach & Maimon, 2005). The similarity defined as the distances between points, however, loses the temporal information in human mobility data (Khoroshevsky & Lerner, 2016; Montoliu et al., 2013; Ye et al., 2009). In our method, we will explore the potential of temporal constraints on optimizing the clustering problem.

We compared our method with state-of-the-art methods based on three datasets. The evaluation metrics include the granularity and the quality of the extracted POIs. Regarding the POI quality, besides precision and recall, we will introduce another metric: predictability limit (Song, Qu, Blumm, & Barabasi, 2010). This metric indicates to what extent the generated POI sequence can be predicted. Although POI clustering and prediction were often discussed simultaneously in related work (Do, Dousse, Miettinen, & Gatica-Perez, 2015; Ikanovic & Mollgaard, 2017; Khoroshevsky & Lerner, 2016), no prior work studied the impact of different clustering methods on the prediction of the generated POI sequence.

The remainder of this chapter is organized as follows: The next section discusses related works. In Section 4.4, we present the details of the proposed approach, including the definitions of the POI and the POI score, as well as our POI extraction algorithm. Section 4.5 reports the experimental results of the proposed approach compared to four other approaches from related work. Following, in Section 4.6, we discuss the contributions and limitations of this work. The final section concludes the chapter and points out the potential future work.

4.3 Related Work

In the field of POI clustering of human mobility, related work often applies the workflow shown in Figure 4.1: The workflow adopted in related work (Ikanovic & Mollgaard, 2017; Montoliu et al., 2013; Ye et al., 2009). (e.g., (Ikanovic & Mollgaard, 2017; Montoliu et al., 2013; Ye et al., 2009)). It contains two steps: Stay-point detection and POI clustering. In the first step, a stay-point detection algorithm - based on the approach proposed by Ye et al. (Ye et al., 2009) - processes the spatio-temporal data sequentially. A stay-point is defined as a region where an individual spends at least a predefined period of time within a certain distance. Because all data points are sequentially processed (a user may return to a place many times), multiple stay-points generated in the first step may
be closely located and should be in one cluster. Therefore, in the second step, a clustering algorithm is used to obtain POIs.

**Figure 4.1:** The workflow adopted in related work (Ikanovic & Mollgaard, 2017; Montoliu et al., 2013; Ye et al., 2009).

Clustering algorithms require parameters which have to be carefully set; one needs to estimate them based on prior knowledge about the data. When using density-based clustering in human mobility data (Ikanovic & Mollgaard, 2017; Sheng et al., 2016; Ye et al., 2009), two parameters should be set: the minimum number of data points in a cluster and the reachability distance threshold (Ester et al., 1996). These parameters reflect how a POI is defined. For example, Ye et al. (Ye et al., 2009) used a set of parameters to define a POI as a geographic area where a) the user stayed more than 30 minutes per visit, b) the user visited more than four times during the user’s overall mobility history, and c) any distance between two data points is less than 200 meters within the geographic area. However, the reason for choosing these values was not explained. Researchers typically set the parameters based on their own experience (e.g., (Montoliu et al., 2013)). In other cases, they did not report how parameters were set at all (Sheng et al., 2016). A value of 150-250 (meters) is often used as the distance parameter in related work (see Table 4.1). However, how these values affect the clustering results is unknown.

**Table 4.1:** Summary of parameters influencing POI definitions in the related approaches. The dashes depict unused parameters in the corresponding approaches. Categories: blue: stay-point detection; orange: POI clustering; green: density-based clustering; gray: hierarchical clustering.

<table>
<thead>
<tr>
<th>Method</th>
<th>Minimum duration $\Delta$</th>
<th>Maximum distance $\theta$</th>
<th>Reachability distance $\varepsilon$ (DBSCAN)</th>
<th>Minimum points $MinPts$ (DBSCAN)</th>
<th>Visit frequency</th>
<th>Duration of visit</th>
<th>Cluster number optimization method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ye et al., 2009)</td>
<td>30 minutes</td>
<td>200 meters</td>
<td>150 meters</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Sheng et al., 2016)</td>
<td>-</td>
<td>-</td>
<td>Not reported</td>
<td>Not reported</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Montoliu et al., 2013)</td>
<td>30 minutes</td>
<td>250 meters</td>
<td>250 meters</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(Khoroshkevsky &amp; Lerner, 2016)</td>
<td>20 minutes</td>
<td>50 meters</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>SC, DB, semantic similarity</td>
</tr>
<tr>
<td>Our Method</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$F_{vd}$</td>
<td>$D_{vd}$</td>
<td>POI score</td>
<td></td>
</tr>
</tbody>
</table>
We conclude the parameters used in the mentioned related work in Table 4.1. The minimum duration in stay-point detection and the minimum points in POI clustering (DBSCAN) are related to the temporal information. However, how these parameters relate to human routines (e.g., daily or weekly visits) is again not addressed in the related work.

The challenge of choosing the appropriate parameters is also pointed out in (Khoroshevsky & Lerner, 2016), where the authors used hierarchical clustering to extract POIs. To get the optimal cluster number (the only parameter in hierarchical clustering), the authors compared methods based on four metrics: Silhouette coefficient (SC) (Rousseeuw, 1987), Davies-Bouldin index (DB) (Davies & Bouldin, 1979), and other two metrics which mix SC and DB with their proposed semantic score. However, temporal information was ignored in the clustering process. Additionally, the hierarchical structure in human mobility environment was not considered. For example, a relatively large POI (e.g., a university) contains several smaller POIs (e.g., the library and the canteen of the university).

Therefore, we design our approach based on three considerations. The approach should:

1. involve both spatial and temporal information for clustering;
2. gain the best granularity based on sophisticated metrics under consideration of the hierarchical structure among POIs;
3. make the parameter setting easier than the prior methods.

### 4.4 Approach for Extracting POIs

In this section, we will discuss the workflow in our POI extraction approach, namely POI clustering with temporal constraints (PC-TC), as shown in Figure 4.2. Our approach consists of three steps. The first step (data pre-processing) will be discussed in the next section, because it depends on the properties of the particular dataset. This section presents the temporal information of human mobility that we will use in PC-TC, the definitions of POI and the POI score, and our POI extraction algorithm.

![Figure 4.2](#)

**Figure 4.2:** The workflow of our approach. It contains three steps: data pre-processing, global POIs clustering, and local POIs clustering.

#### 4.4.1 Temporal Information in Human Mobility

Human mobility, especially for office workers or students, usually follows stable and periodic schedules. The 2D map in Figure 4.3 shows the visited locations and the transitions of two students at Dartmouth College (R. Wang et al., 2014). We
can recognize some obvious clusters (e.g., the ones represented as black circles). Meanwhile, we find it difficult to cluster some areas with sparse location data exclusively based on distance (e.g., the one in the black rectangle). However, in the 3D map in Figure 4.3, it becomes easier to identify the corresponding cluster by adding the time axis. The cluster marked by the rectangle was more frequently visited than other clusters in the surrounding region. This example shows that the temporal information is crucial to understanding and analyzing human mobility patterns, especially for detecting POIs.

Figure 4.3: Location data points and transitions of two college students. In the 2D map (top) the black dots represent data points of location, and the lines show the transitions. The black circles demonstrate two spatially distinct clusters while the black rectangle highlights a potential cluster that cannot be recognized using only spatial information. By adding the time dimension, the black rectangle in the 3D figure (bottom) shows the significant visit frequency of the location in the black rectangle in the 2D map.
4.4.2 Definitions of POIs

Based on related work (Montoliu et al., 2013), we define POIs by considering the visit duration, the visit frequency, and the spatial distance. In our approach, a POI (place of interest) refers to a geographic area where a person (1) stays for a period of time longer than a threshold per visit day on average ($THOLD_D$, e.g., 30 minutes per visit day) and (2) visits a number of days above a threshold ($THOLD_F$, e.g., 3 days per week). Instead of the spatial distance, our POIs are defined by temporal constraints (visit duration and visit frequency). We emphasize the visit frequency and the duration per visit day instead of the overall visit frequency and durations in order to filter out the “occasional visit” noise. For example, a person may visit some places several times but only on a particular day in case of accidents or emergencies. Although we use the spatial distance in our POI clustering algorithm, we do constrain it in our POI definition. Thus, we can achieve the best spatial granularity.

Our daily mobility contains a hierarchical structure. The number of a person’s POIs in real life is small (i.e., several to dozens), generally including home, workplace or school, shopping locations (e.g., malls or supermarkets), and several other places (e.g., cinema and hospital). Some POIs, especially workplace or school, where people spend most of their waking time, may contain multiple smaller significant places such as lab, office, canteen, library, or even a rest place in the garden the user likes to visit. In order to meet our requirement of keeping the hierarchical structure among POIs, we define two types of POIs - global POIs and local POIs - in two tiers as shown in the POI tree in Figure 4.4. One global POI can contain several local POIs. For example, Max goes to the university (global POI) to work, where he visits his office (local POI), the laboratory (local POI), and the campus coffee shop (local POI) on weekdays. There could also be global POIs containing no local POI (e.g., global POI 2 in Figure 4.4). Other clusters in the layer of global POIs refer to the clusters not meeting the requirement of global POIs. However, these clusters may contain special local POIs without belonging to any global POI.

![Figure 4.4: The relationship between global POIs and local POIs.](image)

We think detecting global POIs is important for the following reasons: (1) visits to a global POI should be more frequent than the subordinate local POIs, which may provide more predictability, and (2) detecting global POIs can enable more
in-situ interventions (e.g., providing route suggestions among local POIs to cover more walking if a user is approaching the corresponding global POI).

To sum up, the temporal constraints for POIs include the frequency of visit days \( (F_{vd}) \) and the average duration per visit day \( (D_{vd}) \). Now we show the formal definitions of global and local POIs with the corresponding thresholds for every cluster \( c \) (\( THOLD_{Fg} / THOLD_{Fl} \) refers to the frequency threshold for global/local POIs, while \( THOLD_{Dg} / THOLD_{Dl} \) refers to the duration threshold for global/local POIs, respectively):

\[
\forall c \in \{\text{global POIs}\}, F_{vd}(c) \geq THOLD_{Fg} \quad \text{and} \quad D_{vd}(c) \geq THOLD_{Dg} \quad (4.1)
\]

\[
\forall c \in \{\text{local POIs}\}, F_{vd}(c) \geq THOLD_{Fl} \quad \text{and} \quad D_{vd}(c) \geq THOLD_{Dl} \quad (4.2)
\]

### 4.4.3 POI Clustering Algorithm

Based on the POI definition we now show the details of our POI clustering algorithm. To obtain the optimal spatial granularity, we choose hierarchical clustering as the foundation of our approach. Another reason for choosing hierarchical clustering is that it allows the maintenance of the hierarchical structure of clusters. Thus, we can detect global POIs and local POIs in hierarchical order. When adopting hierarchical clustering, we need to find a metric to determine the optimal cluster number. Instead of using spatial distance-based metrics (Khoroshevsky & Lerner, 2016), we use a new metric called **POI score**, which is the number of POIs among the given clusters. A greater POI score means more extracted POIs which meet our constraints. To find the optimal cluster number, we try all possible cluster numbers, and the one generating the most POIs wins.

![Diagram](image)

**Figure 4.5**: Assume the minimum distance between the two potential POIs in two buildings is 45 meters, then a stay-point detection or DBSCAN with a distance threshold of 50 meters cannot divide the two clusters. The green color indicates locations belong to one POI resulted from DBSCAN.

To illustrate why our hierarchical clustering-based method is superior to the density-based method, we use an example with real-world locations as shown in **Figure 4.5**. We assume that the minimum distance between two potential POIs in two buildings is 45 meters. If DBSCAN is used to cluster the shown data with
a distance threshold of 50 meters, the locations will be in one cluster. Using our approach, we check if more POIs can be generated when we divide them into two clusters or more according to their hierarchical order. In this example, we finally get two POIs (marked with the red or black area).

We conclude the process of finding the optimal cluster number for the hierarchical clustering based on our POI score in Formula 4.3. The function $POI\_SCORE(C_n, P)$ is to calculate the POI score given the POI parameters $P$ and the clusters $C_n$, where $n$ ranges from 2 to the data size $N$.

$$\arg \max_{2 \leq n \leq N} POI\_SCORE(C_n, P)$$

subject to: POI Definition (See Formula 4.1 and 4.2)

**Algorithm 4.1: Extracting Global POIs**

INPUT: The trajectory T; Thresholds THOLD$_g$ and THOLD$_l$ for global and local POIs, respectively.

OUTPUT: The set of POIs G.

1: $G \leftarrow \{\}$;
2: $L \leftarrow \text{LINKAGE}(T)$;
3: $\text{Score} \leftarrow \{\}$;
4: FOR $n = 2 : N$
5: $C_n \leftarrow \text{CLUSTER}(L, n)$;
6: $\text{Score}_n \leftarrow \text{POI\_SCORE}(C_n, \text{THOLD}_g)$;
7: IF MeetTerminateConditions
8: \hspace{1em} BREAK;
9: END
10: END
11: BestClusterNumber $\leftarrow \text{ARG MAX}(\text{Score})$;
12: $C \leftarrow \text{CLUSTER}(L, \text{BestClusterNumber})$;
13: $G \leftarrow \text{POI\_SCORE}(C, \text{THOLD}_g)$;
14: $G \leftarrow G \cup \text{POI\_SCORE}(C, \text{THOLD}_l)$;
15: RETURN $G$

**Algorithm 4.1** demonstrates the procedure of extracting global POIs. Given the spatio-temporal data of human mobility and the POI parameters, the hierarchical clustering (C-Linkage (Defays, 1977)) is firstly applied (Algorithm 4.1, line 2). Next, the clustering results with cluster number $n$ from 2 to $N$ are evaluated using the POI score. Normally, the POI score increases at the beginning of the iterations when the cluster number is smaller than the actual maximum of POI score; it decreases when the cluster number becomes too large that fewer clusters can meet the POI constraints. In the extreme case that every data point represents a cluster, it is likely that no cluster could meet the temporal constraints. Therefore, we define some conditions where the loop can be terminated to avoid unnecessary computation. These conditions include: (1) the POI score equals zero; (2) the POI score does not increase for more than 50 iterations (large enough); (3) the POI score decreases and becomes smaller than the highest by 10
4.4 Approach for Extracting POIs

(large enough). This strategy saves computational resources when the data size is large.

We use the obtained optimal cluster number to generate the POIs. The clusters that meet the constraints of global POI will be used for clustering local POIs, while the other clusters will be checked if they meet the constraints of local POI (Algorithm 4.1, line 14). The process for local POIs is the same but with the constraints for local POI (Algorithm 4.1, line 6).

The worst-case computation complexity of searching for the maximum POI score is $O(n^2)$. As the computation complexity of the adopted hierarchical clustering algorithm (Algorithm 4.1, line 2) is $O(n^2)$, the total complexity of Algorithm 4.1 is $O(n^2)$. To improve the computation time, we propose a distance computation method for hierarchical clustering.

4.4.4 Distance Computation in Hierarchical Clustering

In our evaluations, we use a simpler distance calculation method instead of the Haversine formula in hierarchical clustering. The Haversine formula is often used to estimate the distance between two GPS locations on earth given the latitude values $\phi$ and longitude values $\lambda$, as shown in Formula 4.4. The variable $d$ represents the distance between the two points, and $r$ is the Earth’s radius.

$$hav\left(\frac{d}{r}\right) = hav(\phi_1 - \phi_2) + \cos \phi_1 \cos \phi_2 \, hav(\lambda_1 - \lambda_2) \tag{4.4}$$

Where: $$hav(\theta) = \sin^2 \left(\frac{\theta}{2}\right) = \frac{1 - \cos \theta}{2}$$

To get the distance $d$ by solving this equation is computationally expensive. However, it is not necessary to calculate the absolute distances on earth in hierarchical clustering. Instead, the relative difference determines the hierarchical structure. Assume that the distances between points A and B (i.e., $d_1$) and points C and D (i.e., $d_2$) should be compared on a sphere, with $d'$ as the Euclidean distance between two points on a sphere. Given that $\alpha_1$ and $\alpha_2$ are the radians of arc AB and arc CD, respectively, we can calculate $d'$ and $d$ using Formula 4.5 and 4.6. From THEOREM 1, we see that the Euclidean distance can replace arc length to compare distances.

$$d = ar \tag{4.5}$$

$$d' = 2r \sin \left(\frac{\alpha}{2}\right) \tag{4.6}$$

\text{THEOREM 1:} \quad \text{if } d_1 > d_2, \text{ then } d'_1 > d'_2

\text{PROOF:} \quad \text{if } d_1 > d_2, \text{ then } \alpha_1 > \alpha_2,$n

then $\sin \left(\frac{\alpha_1}{2}\right) > \sin \left(\frac{\alpha_2}{2}\right)$ when $0 < \alpha < \pi$

then $d'_1 > d'_2$
In other words, we can use the Euclidean distance in hierarchical clustering. It should be noted that the values of latitude and longitude cannot be used directly as in 2D Cartesian coordinates, because a distance represented by unit degree is not consistent with the latitude direction. The question is thus how to acquire the Euclidean distance using the latitude and longitude value. Instead of mapping points from spherical coordinates to Cartesian coordinates to calculate the Euclidean distance, we introduce an approximate distance based on equirectangular projection (Amoruso, 2016) as shown in Formula 7. In this formula, $x$ is the parameter to allow different spatial metrics, e.g., $x=1$ means Manhattan distance and $x=2$ means Euclidean distance. The key in this formula is to adjust longitude values to make the distance indicated by the unit degree of any longitude value be the same as of the latitude value. This approximate distance is not as accurate as the distance calculated by the Haversine formula. However, the accuracy is acceptable when clustering human mobility where the distances are relatively small.

$$d = [ |\varphi_1 - \varphi_2|^x + |\lambda_1 \cos \varphi_1 - \lambda_2 \cos \varphi_2|^x ]^{1/x} \quad (4.7)$$

The computation components for distances in a sequence with $n$ points using the Haversine formula and our method ($x=1$) are listed in Table 4.2, where we can see that our method can save considerable computing time compared to using the Haversine formula.

Table 4.2: Computation load comparison between Haversine formula and the simplified method. $C^n_2$ refers to the number of combinations of 2 elements out of the total elements $n$. Here $C^n_2$ indicates the number of distances between any two locations in all data points.

<table>
<thead>
<tr>
<th></th>
<th>Haversine Formula</th>
<th>Our Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosine</td>
<td>$n + 2C^n_2$</td>
<td>$n$</td>
</tr>
<tr>
<td>Arccosine</td>
<td>$C^n_2$</td>
<td>0</td>
</tr>
<tr>
<td>Multiplication</td>
<td>$3C^n_2$</td>
<td>$n$</td>
</tr>
<tr>
<td>Plus</td>
<td>$4C^n_2$</td>
<td>$3C^n_2$</td>
</tr>
<tr>
<td>Modulus</td>
<td>0</td>
<td>$2C^n_2$</td>
</tr>
</tbody>
</table>

4.5 Evaluation

4.5.1 Datasets

We used three datasets (two public and one private) to evaluate our method. The first dataset is from the StudentLife Study (R. Wang et al., 2014) which contains heterogeneous data of 49 students collected from a class of Dartmouth College in the U.S. for around ten weeks in 2013. The GPS location data was collected every 10 or 20 minutes by a dedicated mobile application on participants’ Android smartphones. The other dataset comes from the Mobile Data Challenge (MDC) (Kiukkonen, Blom, Dousse, Gatica-Perez, & Laurila, 2010; Laurila et al., 2012),
which was collected from Oct. 2009 to Mar. 2011 and involved around 200 people in Switzerland. The data was collected by a dedicated mobile application in Nokia N95; most of the participants were employees or college students. Unlike the first dataset, the time interval of data collection is from 1 minute to 10 minutes according to the state of the phone (Kiukkonen et al., 2010). We show more details of the two datasets in Table 4.3.

Table 4.3: Details about the StudentLife dataset and the MDC dataset.

<table>
<thead>
<tr>
<th></th>
<th>MDC Dataset</th>
<th>StudentLife Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td>October 2009 - March 2011</td>
<td>Spring term in 2013 (10 weeks)</td>
</tr>
<tr>
<td><strong>Device</strong></td>
<td>Nokia N95</td>
<td>Android smartphone</td>
</tr>
<tr>
<td><strong>Users size</strong></td>
<td>191 participants</td>
<td>49 college students</td>
</tr>
<tr>
<td></td>
<td>most between the ages of 22-27(30%), 28-33(37%) years old</td>
<td>undergraduates and graduate students</td>
</tr>
<tr>
<td></td>
<td>63% of the users were employed, 26% were students</td>
<td></td>
</tr>
<tr>
<td><strong>Data collection scheme</strong></td>
<td>state-based (1-10 min)</td>
<td>time window-based (10/20 min.)</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>GPS 2.4%, WLAN 54.6%, GSM 43%</td>
<td>GPS (100%)</td>
</tr>
</tbody>
</table>

The data collection duration in the MDC dataset (18 months) is much longer than the one in the StudentLife dataset (10 weeks). To make the MDC dataset comparable to the StudentLife dataset, we exclusively selected the participants’ data from the MDC dataset covering more than 60 days, which resulted in a dataset containing the data of 144 participants. For this dataset, we only keep 60 days’ data for each participant.

Besides the two public datasets, we also tested our method in a mobile health study based on a mobile application with 16 participants (see details in Chapter 6). The mobile application logged the users’ GPS location every five seconds when the phone is moving or every 20 minutes when it is still. We used one-week data during the study to evaluate our proposed method.

4.5.2 Data Pre-processing

Since all human mobility data were collected during the users’ real-life activities, erroneous data or missing data are inevitable. Any missing data must be detected in the pre-processing step because it is related to the duration estimation. Since all used datasets have a minimum time interval for data collection, a time interval of 30 minutes was used to identify missing data. We regard the location data with a confidence radius larger than 1,000 meters or equal to 0 as erroneous data.

After labeling the missing data and removing the error data, we deleted the continuously un-changed location data and retained the first valid location data. Then we calculated the stay durations of each location point.
4.5.3 Compared Approaches and Parameters

From our review of related work, we selected four other approaches to compare with our approach (see Table 4.4). All compared approaches follow the workflow shown in Figure 4.1. Approach 1 and 2 use density-based clustering (OPTICS (Ankerst, Ankerst, Breunig, Kriegel, & Sander, 1999) and DBSCAN (Ester et al., 1996)). Compared to DBSCAN, OPTICS only needs one parameter (the minimum point number MinPts) (Sander et al., 2003). The implementation of the OPTICS approach is adapted from (Pavlopoulos, Moschopoulos, Hooper, Schneider, & Kossida, 2009), while the DBSCAN approach is implemented based on (Yarpiz, 2018). Approach 3 and 4 use hierarchical clustering algorithms (complete-linkage or C-Linkage) (Defays, 1977), where Davies-Bouldin (DB) index (Davies & Bouldin, 1979) and Silhouette coefficient (SC) (Rousseeuw, 1987) are used to evaluate the generated clusters and find the optimal number of clusters, respectively. The built-in functions of C-Linkage, DB, and SC in Matlab are used in corresponding approaches. In the stay-point detection, we adopt the algorithm described in (Ye et al., 2009), with one improvement - we considered the missing data when calculated the stay-points. We avoided accumulating the duration of missing data points, which make the stay-point detection be more accurate. Please refer to our source code in the supplementary material for more details.

Table 4.4: The list of different approaches and parameters in the experiments. SP means stay-point detection. The N in MinPts means the data size. The other symbols have the same meaning as in Table 4.1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Approach</th>
<th>Abbr.</th>
<th>Δ (minutes)</th>
<th>θ (meters)</th>
<th>ε (meters)</th>
<th>MinPts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SP + OPTICS</td>
<td>OPTICS</td>
<td>30</td>
<td>50</td>
<td>-</td>
<td>(\log_{10}(N))</td>
</tr>
<tr>
<td>2</td>
<td>SP + DBSCAN</td>
<td>DBSCAN</td>
<td>30</td>
<td>50</td>
<td>50</td>
<td>(\log_{10}(N))</td>
</tr>
<tr>
<td>3</td>
<td>SP + C-Linkage(DB)</td>
<td>DB</td>
<td>30</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>SP + C-Linkage(SC)</td>
<td>SC</td>
<td>30</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>PC-TC (Ours)</td>
<td>PC-TC</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

We used a technique (see the previous section) to simplify the distance computation using GPS data when the absolute distance is not necessary. Among the five approaches, only DBSCAN needs absolute distances except for stay-point detection. Therefore, we applied the technique to all the other approaches.

We conducted three evaluations using different parameter settings. First, we use all the listed methods with a set of heuristic parameter values as shown in Table 4.4. Second, we only compare our method with DBSCAN by choosing several values for the key parameters in a reasonable range. Third, we deployed our method in a mobile application for a user study about sedentary behavior change.
All tested approaches were implemented in Matlab 2017a, and the evaluation was run on a Lenovo ThinkPad laptop with 8 GB memory and Intel i7-5600U CPU (2.6 GHz). In the third evaluation, we implemented our method in Java and tested it on Android smartphones.

4.5.3.1 First Evaluation

We use 50 meters (Montoliu et al., 2013; Ye et al., 2009) and 30 minutes (Khoroshevsky & Lerner, 2016) as the corresponding parameters in stay-point detection. The parameter MinPts is set to $\log_{10}(N)$ as recommended in (Birant & Kut, 2007) in OPTICS and DBSCAN. The distance parameter in DBSCAN is also set to 50 meters, the same as the distance parameter in stay-point detection. In our approach we set the parameters for the office working scenario as following. For a global POI (e.g., workplace), we assume that a user visits the location every day during weekdays. Thus, the threshold for $F_{vd}$ should be around $5/7$. Considering an exception rate of 10% (e.g., traveling somewhere else), we suggest using 0.63 as the threshold for $F_{vd}$ of global POIs. Similarly, we assume that a user visits a local POI at least once a week and suggest using 0.13 (1/7 * 90%) for local POIs. In terms of the $D_{vd}$, we use 120 minutes as the threshold for global POIs, and 30 minutes for local POIs, the same as in stay-point detection.

4.5.3.2 Second Evaluation

To investigate the effect of parameters in PC-TC, we need to conduct another evaluation. We only compare our approach with DBSCAN, which is the most used approach in related works. In TC-PC, we set $F_{vd}$ of local POIs to a range of values (0.1-0.9). These values cover visit patterns from weekly to daily. In DBSCAN, we use 4 values (1, 5, 25, and 50 meters) for the distance parameter. The other parameters remain the values in the first evaluation.

4.5.3.3 Third Evaluation

To evaluate how PC-TC works in real-world applications, we implemented it in a mobile application for a mobile health study. Unlike the previous evaluations, this one allows us to test the accuracy of the generated POIs by asking the users to name the POIs. The mobile application monitored users’ steps and location in real-time; PC-TC detects the places (POIs) where a user is sedentary for more than one hour. Therefore, we only detected local POIs with $F_{vd} = 0.1$ and $D_{vd} = 60$ (minutes). This study lasted three weeks in November 2018 and involved 16 participants from the local university.

4.5.4 Evaluation Metrics

In the first two evaluations, we compared the chosen approaches based on three metrics: the number of POIs, the predictability limit (Ikanovic & Mollgaard, 2017; Song et al., 2010) of POI sequence, and the computation time. The number of POIs indicates how much information or patterns are extracted; the predictability limit
Chapter 4: Understanding Users’ Mobility Patterns: Clustering

shows the quality of the POIs regarding next-place prediction; the computation
time reflects the approaches’ efficiency.

The number of POIs and the computation time are intuitive metrics, but the
predictability limit is relatively difficult to understand and not well used. So it is
necessary to introduce its concept and applications.

The upper limit of human mobility predictability (predictability limit, PL)
represents the average probability of correctly predict the next POI by a proper
algorithm, given a POI sequence. Song et al. (Song et al., 2010) initially
investigated the PL using cellular-scale location data of cellphone as human’s
mobility data. The PL is related to the entropy rate $S$ of a given sequence of
length $N$, which can be estimated using the Lempel-Ziv compression algorithm
as shown in Formula 4.8, where $N$ is the length of the location sequence and $A_i$
refers to the length of the shortest substring starting at position $i$ which does
previously not appear from position 1 to $i - 1$. It should be noted that the PL is
only a theoretical value while it is unclear how to approach the limit.

Nevertheless, we can still use it as a metric to compare sequences in the
perspective of their potentials of being predicted.

\[
S = \left( \frac{1}{N} \sum_{i=1}^{N} \frac{A_i}{\log(N)} \right)^{-1}
\]  

(4.8)

Naturally, a larger number of POIs in a sequence provides more mobility
information, which can potentially increase the prediction uncertainty and
decrease the PL (Ikanovic & Mollgaard, 2017). This is in line with our intuition
that more information might make it more difficult to predict. However, from
Formula 4.8, we know that the gain of the entropy rate is caused by new
subsequences. It means infrequent POIs in a sequence will increase the prediction
difficulty and decrease the PL. Our clustering method intends to obtain the most
POIs with temporal constraints, which can simultaneously avoid places below a
frequency threshold (e.g., a restaurant visited only once). Thus, it may decrease
the entropy rate and increase the predictability limit. We want to explore how
our method can balance this tradeoff in comparison to other methods. To the best
of our knowledge, this is the first time the PL is used to evaluate POI clustering
results.

In the third evaluation, as we could get the ground-truth from the participants in
our study, we used precision and recall to show the performance of our method.

4.5.5 Results

4.5.5.1 First Evaluation

The results of the first evaluation are shown in Figure 4.6. The compared
approaches define no such temporal constraints for POI-determination as we
define in our approach. Thus, we regard all generated clusters in the other
approaches as POIs. The numbers of POIs are shown in the two leftmost graphs
in Figure 4.6. For the StudentLife dataset, our approach (PC-TC) generates significantly more POIs than other approaches, while for the MDC dataset our approach yields less POIs than DB and SC.

The StudentLife dataset only contains data from students of Dartmouth College, which is a self-contained campus (R. Wang et al., 2014). Therefore, most of the participants’ location data were generated on campus. The potential POIs, such as the dormitory, the classrooms, and the library, are close to each other. Referring to the example in Figure 4.5, POIs with a distance of fewer than 50 meters (the threshold for stay-points) will not be separated in the compared approaches. In PC-TC, there is no constraint for spatial distance, which is why PC-TC obtains more POIs. Unlike the dataset representative of campus life, the visited places of the participants in the MDC datasets are generally farther apart from each other. We observed that many participants in this dataset visited several cities with low frequency in Switzerland. In PC-TC, the places which cannot meet the temporal constraints are filtered out, which explains why the numbers of POIs in PC-TC for the MDC dataset do not exceed the ones of other approaches as for the StudentLife dataset. By contrast, DB and SC generate much more POIs because they have no constraints on visit duration or frequency of POIs.

In addition to the total number of local POIs in PC-TC, we also exclusively show the number of global POIs. The numbers of global POIs vary from one to five for all participants of both datasets, which is reasonable in the real world.

As discussed the previous section, there is a trade-off between POI number and predictability limit (PL). The two graphs the center of Figure 4.6 show that for the StudentLife dataset our approach generates more POIs but results in a lower PL (0.67 in the median). However, with much larger POI numbers, our approach obtains nearly the same PL as OPTICS (0.71 in the median) for the StudentLife dataset. Meanwhile, for the MDC dataset, with very similar POI numbers (6-10 in the median), our approach achieves a higher PL (0.92 in the median) than OPTICS (0.76 in the median) and DBSCAN (0.83 in the median). This demonstrates the potential of our approach of balancing the tradeoff between POI number and PL.

Since the numbers of global POIs are very small, the predictability limit levels are always higher than for local POIs. Therefore, global POI sequences provide an alternative in next place prediction tasks.
Figure 4.6: The evaluation results of different approaches for the StudentLife dataset and the MDC dataset. The two leftmost graphs show the numbers of extracted POIs; the two middle graphs present the predictability limit of the POI sequences; the two rightmost graphs represent the computation time.

In terms of computation time, shown in the rightmost graphs in Figure 4.6, our approach requires more time than the other approaches. This result can be explained by two reasons. First, the computation complexity of PC-TC ($O(n^2)$) is higher than DBSCAN or OPTICS ($O(n \log(n))$). Second, the stay-point detection step shrinks down the data point number to feed into the following clustering algorithm. To free the spatial constraint, PC-TC does not use stay-point detection. Therefore, PC-TC has more input data than other approaches. If the compared approaches should achieve a finer spatial granularity, they have to use a smaller distance parameter for the stay-point detection to produce a higher quantity of input data. Consequently, the compared approaches are expected to require more computation time. Our next evaluation confirms this inference.

As should be noted, the computation using approach 3 and 4 spend a very long time when the cluster number is large (e.g., several thousand). Such large cluster numbers make no sense in practice because human mobility is limited and one participant from the dataset hardly have visited thousands of places in two months. As the data sizes of the selected datasets after pre-processing are 3,648 and 3,984 data points on average respectively, we limit the cluster number to half of the input data size for DB and SC.
4.5.5.2 Second Evaluation

Figure 4.7 shows the results of our second evaluation. For the StudentLife dataset, the POI numbers by PC-TC decrease in a linear trend with the threshold values of $F_{vd}$ increasing from 0.1 to 0.9. Although the values of $\varepsilon$ (5, 25, 50 meters) can also moderate the POI number by DBSCAN, the effect is small. As expected, the corresponding PLs change in the opposite direction. Interestingly enough, when PC-TC ($F_{vd}=0.9$) and DBSCAN ($\varepsilon=5$) produce the comparable level of POI number, the PL (0.82 in the median) by PC-TC is significantly higher than the one (0.77 in the median) by DBSCAN. We also find the same phenomena for the MDC dataset, which indicates that PC-TC has more potential for next place prediction applications. Regarding the computation time, DBSCAN requires more time when the value of $\varepsilon$ becomes smaller so that the stay-point detection produces more input data for DBSCAN.

Figure 4.7: The results of the second evaluation. The two leftmost graphs show the numbers of extracted POIs; the two middle graphs present the predictability limit of the POI sequences; the two rightmost graphs represent the computation time.

4.5.5.3 Third Evaluation

We could not evaluate the accuracy of the detected POIs in the previous evaluations. Only the data owner can confirm if a detected POI is a meaningful place (e.g., a user often goes to a place near his office to smoke) or if there is any POI the algorithm cannot detect. In the last evaluation, therefore, we validated our POI clustering results along with a mobile health study focusing on reducing users’ prolonged sedentary behavior (see Chapter 6 for details). By showing the detected POIs based on one-week data for each participant on the map in our app, we asked them to name the detected POIs one by one. The ones the
participants can recognize (e.g., the office and the library) were labeled the correct POIs. Then we asked the participants to recall if there was any other place where they spent much time but was not shown on the map. These were labeled as the missed POIs.

Table 4.5: The results of the third evaluation.

<table>
<thead>
<tr>
<th>User</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected POI</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Correct POI</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Missed POI</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The evaluation results are shown in Table 4.5. Among all the detected POIs, only user #6 and user #7 had one POI that they could not recognize. Regarding the missed POI, only user #9 mentioned one place she stayed for much time (about two hours) that did not show on the map. While our algorithm could separate close but different places (e.g., the office and the laboratory, or different zones in the library) – as we illustrated in Figure 4.5, we also found it is sensitive to the noise in GPS data. In Figure 4.8, the dots were the GPS locations. Only the dots in the middle represented the true location; the others were not accurate because of the unstable GPS signal. The triangles represented the multiple detected POIs, which were caused by the noise of GPS data.

![Image of GPS data with multiple POIs](image)

Figure 4.8: The multiple POIs caused by the noise in GPS data.

4.6 Discussion

In comparison with other approaches, PC-TC uses temporal constraints instead of spatial constraints for POI extraction. The parameters required to define a POI are more intuitive (according to our routines) and can be flexibly adjusted in practice. Although PC-TC also needs heuristic parameters (visit duration and frequency), it shows some superior properties. Our evaluation results show that PC-TC performs better than compared approaches regarding the number of extracted POIs and the predictability limit of the POI sequences. In other words, with the same level of POI numbers, the extracted POI sequences by PC-TC achieve higher predictability limits. Although we can reduce its computation
time using our distance computation method, the complexity of PC-TC is $O(n^2)$. This will limit its usage on large datasets, but it is still applicable in many real-world applications.

By following a new optimization objective, POI score, our approach has demonstrated its potential for next place prediction. Theoretically, the POI score can be applied to any hierarchical clustering method. We use the complete-linkage algorithm in our approach as in (Khoroshevsky & Lerner, 2016). However, we did not test other hierarchical clustering algorithms, which is one limitation of this work.

A second limitation applies to the datasets we used in our evaluation. The GPS location data in the StudentLife dataset were collected only every 10 or 20 minutes, which may miss mobility information and decreases the accuracy of duration estimation. The GPS location data in the MDC dataset only covers 2.4% of participants’ time during the data collection period. We suggest that researchers working on data collection in this domain could use state-based methods (Kiukkonen et al., 2010; Y. Wang, Duan, et al., 2016), which collects data points based on the change of a user’s state instead of a time interval, to avoid losing much mobility information of users and to reduce the redundancy in the dataset.

4.7 Conclusion

In this chapter, we proposed an approach, POI clustering with temporal constraints (PC-TC), to extract places of interest using spatio-temporal data of human mobility. Taking advantage of the hierarchical information embedded in human mobility data, PC-TC can extract POIs in two scales (i.e., global POIs and local POIs). Instead of using spatial parameters, PC-TC explores the potentials of temporal constraints in human mobility POI clustering. Our evaluation results based on two public datasets showed that PC-TC could extract more information from human mobility data without losing the prediction potential. The results of our field study applying the proposed method in a mobile application suggest that PC-TC is capable of serving in real-world applications.

Potential future work includes: (1) investigating the impact of different parameter values on POI extraction results, (2) comparing different linkage algorithms (e.g., Ward’s method (Ward, 1963)) in our approach, and (3) exploring online hierarchical algorithms methods (e.g., BIRCH (T. Zhang, Ramakrishnan, & Livny, 1996)) to allow long-term online analysis.

We have used the PC-TC approach in our field study, in which the we extracted and visualized users’ sedentary places (POIs) to support users’ action planning for sedentary behavior change (see Chapter 6 for details).
5 UNDERSTANDING USERS’ MOBILITY PATTERNS: PREDICTION

"因地制宜"; 因时制宜。 Adapting measures according to local conditions; Adapting measures according to the time change.

- 1《淮南子》
- 2《吴越春秋》

1 Huainanzi
2 Spring and Autumn Annals of Wu and Yue
(Ancient Chinese texts)


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Corinna Breitinger and Björn Sommer helped in discussing the data analysis, and reviewing the paper. Falk Schreiber and Harald Reiterer supervised the work and reviewed this paper.
5.1 Abstract

In the domain of human behavior prediction, next-place prediction is an active research field. While prior work has applied sequential and temporal patterns for next-place prediction, no work has yet studied the prediction performance of combining sequential with temporal patterns compared to using them separately. In this chapter, we address next-place prediction using the sequential and temporal patterns embedded in human mobility data that has been collected using the GPS sensor of smartphones. We test five next-place prediction methods, including single pattern-based methods and hybrid methods that combine temporal and sequential patterns. Instead of only examining average accuracy as in related work, we additionally evaluate the selected methods using incremental-prediction accuracy on two publicly available datasets (the MDC dataset and the StudentLife dataset). Our results suggest that (1) integrating multiple patterns is not necessarily more effective than using single patterns in average prediction accuracy, (2) most of the tested methods can outperform others for a certain time period (either for the prediction of all places or each place individually), and (3) average prediction accuracies of the top-three candidates using sequential patterns are relatively high (up to 0.77 and 0.91 in the median for both datasets). For real-time applications, we recommend applying multiple methods in parallel and choosing the prediction of the best method according to incremental-prediction accuracy. Lastly, we present an expert tool for visualizing the prediction results.

5.2 Introduction

In many mobile health (WHO, 2011) applications, recommendations should be delivered to users before the target behavior occurs. Therefore, human behavior prediction plays a very important role. For example, Rahman et al. predicted the “about-to-eat” event to design food intake interventions. In the research field of just-in-time adaptive interventions (JITAls), the prediction of the opportune moments (Poppinga, 2014) or meaningful moments (Y. Wang, Duan, et al., 2016) determines both the effectiveness of the interventions and the user experience. Among human behavior prediction studies, location-based applications are drawing much attention, since human behaviors are usually coupled with specific places and the corresponding information is relatively easy to collect using smartphones (Naughton et al., 2016; Rabbi et al., 2015).

In the domain of human behavior prediction, next-place prediction is currently an active research field (Do & Gatica-Perez, 2014; Khoroshevsky & Lerner, 2016; Liu, Wu, Wang, & Tan, 2016; Z. Zhang et al., 2017). To our knowledge, there is no study applying next-place prediction to mobile health applications. We see high potential of using next-place prediction to provide personalized and adaptive mobile health interventions and recommendations. Next-place prediction can be regarded as a contextual prediction problem where the next place
is assumed to depend on a user’s context of current and previous states (Do et al., 2015). The context refers to any information related to the prediction (e.g., the user’s routine, calendar events, weather forecast). However, context information is not always available in practice due to privacy concerns or other limitations (e.g., manual logging). In this work, we focus on sequential and temporal patterns that can be derived directly from the past trajectories of human mobility.

Sequential patterns stand for repeated sub-sequences appearing in a specific order within a given sequence, e.g., Alice often (8 out of 10 times) visits place B after place A according to her previous month’s mobility data. Temporal patterns refer to the time distribution for visited places. For instance, Bob arrives at his office between 8:30 am and 9:00 am on Mondays (3 out of 4 times) given his previous month’s mobility data.

In a recent work, Ikanovic and Mollgaard (Ikanovic & Mollgaard, 2017) examined the limits of predictability in human mobility using GPS traces collected by smartphones from 604 individuals in Denmark. The upper limit of predictability for next-place prediction using the method by Song et al. (Song et al., 2010) was 71.1%, while the average prediction accuracy using a first-order Markov chain was 39.8%. Although Song et al. (Song et al., 2010) argue that the upper limit calculated by their method is in principle attainable by an appropriate algorithm, such an algorithm has not been found yet. However, we can still use the predictability limit to compare how far an algorithm’s performance is away from the theoretical limit. In addition to using a first-order Markov chain as in (Ikanovic & Mollgaard, 2017), we compare the performance of a more sophisticated Markov chain-based model, Active LeZi, to the predictability limit using two publicly available datasets, the Mobile Data Challenge (MDC) (Kiukkonen et al., 2010; Laurila et al., 2012) dataset and the StudentLife dataset (R. Wang et al., 2014).

Markov chain-based models only consider sequential patterns. In the methods we evaluate, we also consider a temporal pattern-based method inspired by the work of Khoroshkevsky and Lerner (Khoroshkevsky & Lerner, 2016). Although prior work has already used sequential and temporal patterns for next-place prediction, the performance of combining both sequential and temporal patterns compared to using single patterns has not yet been studied. In our study, we include not only three single pattern-based methods but also two sequential and temporal hybrid methods using joint conditional probability. Thus, our study allows comparing the prediction performance between these methods for the first time.

Another missing component in related work is that only the prediction performance on average among all users’ data has been analyzed. To provide an in-depth analysis on the prediction results at the level of the individual and even at the level of the Places of Interest (POI) over time, we use incremental-prediction
accuracy and develop an interactive visualization tool to present the evaluation results.

In this work, we examine the following research questions:

(Q1) What is the prediction performance of using the selected sequential and temporal patterns separately?

(Q2) What is the prediction performance of using sequential patterns and temporal patterns in combination based on joint conditional probability?

(Q3) Can any method outperform other methods for all of the users’ data on average?

(Q4) Can any method outperform other methods for all places or for one specific place at the level of the individual, as well as at the POI level over time?

The remainder of this chapter is structured as follows. In Section 5.3, we clarify two models of human mobility from the literature and describe the prediction task based on our selected model. Subsequently, we present related work (Section 5.4), before describing our methodology to address the stated research questions (Section 5.5). In Section 5.6, we present the evaluation and report on our results. Then, we introduce our visualization tool for analyzing the results in more detail, followed by a discussion of the contributions and drawbacks of our evaluation. Finally, we conclude the chapter and point out future work.

5.3 Problem Statement

The spatio-temporal sequential data of a person’s mobility trajectory can be illustrated as shown in Figure 5.1. Blocks represent places where a person spends a period of time, blocks with dashed shapes illustrate predicted ones, while blue arrows show transitions between these places. Places are extracted from GPS traces using clustering algorithms as in Chapter 4 (Khoroshevsky & Lerner, 2016; Y. Wang, Sommer, Schreiber, & Reiterer, 2018; Ye et al., 2009) and mapped into symbol sequences as shown in Table 5.1.

Table 5.1: A symbolic-temporal sequence representing human mobility.

<table>
<thead>
<tr>
<th>Place</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrival Time</td>
<td>2017-1-8 9:00</td>
<td>2017-1-8 10:20</td>
<td>2017-1-8 11:10</td>
<td>2017-1-8 11:40</td>
<td>2017-1-8 13:00</td>
</tr>
<tr>
<td>Departure Time</td>
<td>2017-1-8 10:00</td>
<td>2017-1-8 11:00</td>
<td>2017-1-8 11:30</td>
<td>2017-1-8 12:40</td>
<td>2017-1-8 14:00</td>
</tr>
</tbody>
</table>
Figure 5.1: The human mobility model, containing concrete person’s locations (continuous boxes) and predicted ones (dashed boxes).

Given an individual’s historical mobility data and current state (i.e. place and time), the prediction problem thus encompasses the following sub-questions:

1. What is the predicted leaving time from the current place \( t_{\text{leave}} \) in Figure 5.1? 
2. What will be the next visit place \( P_{\text{next}} \) in Figure 5.1? 
3. What will be the arrival time at \( P_{\text{next}} \) \( t_{\text{arrival}} \) in Figure 5.1?

As we can see in Figure 5.1, if a person is currently in transition between two places (Prediction Point 2 in Figure 5.1), only question 2 and 3 should be answered. In this case, next-place prediction is easier because more information is provided (e.g., moving direction). In this chapter, we focus on question 2 at Prediction Point 1 as shown in Figure 5.1.

The data model we adopt is in line with (Ikanovic & Mollgaard, 2017), where the next place is always different from the current one. In comparison, another data model (shown in Table 5.2) represents a time bin-based sequence. We can see that this model contains continuous duplicates of places (Do et al., 2015; Lu, Wetter, Bharti, Tatem, & Bengtsson, 2013; Song et al., 2010). With such a data model, next-place prediction is equivalent to next-bin prediction, which means that each place is computed for the next time bin featuring an individual duration (Ikanovic & Mollgaard, 2017). The prediction accuracy in this model is greatly affected by the size of the time bin. When the time bin size is very small (e.g., one minute), the prediction accuracy will be very high by just using the current place as the prediction of the place in the next time bin. However, this high prediction accuracy only shows that the individual has stayed at the same place. It can be difficult to choose the appropriate size of the time bin in practice. Once the size of the time bin has been chosen, the time scale is fixed, which cannot be flexibly adjusted in practice.

Table 5.2: Time bin-based symbolic-temporal sequence (based on the sequence in Table 5.1).

<table>
<thead>
<tr>
<th>Place</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>B</th>
<th>B</th>
<th>B</th>
<th>C</th>
<th>B</th>
<th>B</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2017-1-8 9:00</td>
<td>9:20</td>
<td>9:40</td>
<td>10:00</td>
<td>10:20</td>
<td>10:40</td>
<td>11:00</td>
<td>11:20</td>
<td>11:40</td>
<td>12:00</td>
</tr>
</tbody>
</table>
5.4 Related Work

5.4.1 Sequential Patterns

The simplest model of sequential patterns is a first-order Markov chain (MC1), which assumes that the prediction of the next place is only related to the current one. This method is often used as a naive baseline method when evaluating other more sophisticated methods (Ikanovic & Mollgaard, 2017; Khoroshevsky & Lerner, 2016; Petzold, Bagci, Trumler, & Ungerer, 2006). We also include this method in our evaluation.

Unlike fixed-order Markov models, such as MC1, Active LeZi is an incremental parsing algorithm that can extract sequential patterns of different lengths (Gopalratnam & Cook, 2007). Stemming from the family of LZ-based data compression algorithms (Bhattacharya & Das, 2002; Rodriguez-Carrion et al., 2012; Ziv & Lempel, 1978), Active LeZi constructs a tree to contain the sequential patterns. Active LeZi solves the problems of ignoring cross-boundary patterns and embedded patterns known from previous LZ-based algorithms (Rodriguez-Carrion et al., 2012). In addition, Active LeZi adopts a partial prediction match (PPM) strategy (Cleary, Teahan, & Witten, 1984) for probability assignment to solve the zero-frequency problem. Another property of Active LeZi is that longer sequential patterns are weighted higher than shorter patterns. The reason is that longer sequential patterns appear later in the sequence, i.e., the more recent patterns have more impact on the current prediction when mapped to next-place prediction. While Active LeZi has shown promising prediction accuracy when applied in the smart home setting (Gopalratnam & Cook, 2007) and for human mobility at the cellular scale (Cheng, Qiao, & Yang, 2017), no prior work has examined Active LeZi when applied to next-place prediction on the more granular GPS data collected by smartphones in a real-world outdoor setting.

5.4.2 Temporal Patterns

Compared to sequential patterns, temporal patterns (e.g., visiting time and day of week) are more intuitive because they can be understood as schedules. For example, Cindy, is an office worker. During weekdays, she goes to work at 9:00 am and leaves her office at 5:00 pm. Khoroshevsky and Lerner (Khoroshevsky & Lerner, 2016) included five temporal pattern-based methods in their evaluation, which use the place: (1) with the highest conditional probability given the day of the week; (2) with a non-zero conditional probability given the day of the week, while being spatially nearest to the current place; (3) with a non-zero conditional probability given the hour of the day and being temporally closest to the current time; (4) with the highest conditional probability given the day of the week and the hour of the day; (5) with a non-zero conditional probability given both the day of the week and the hour of the day, while being spatially the closest to the current place. Surprisingly, the average prediction accuracies (44 - 46.5%) were much lower than when using MC1 (70.25%) (Khoroshevsky & Lerner, 2016). Do
et al. used several temporal patterns including time-of-day (similar to (4)), day-of-week (same as (1)), is-weekend, and temporally close places in their study (Do et al., 2015). However, none of the temporal patterns were tested separately. We include a temporal pattern (see Sect. 4 for details) in our evaluation based on (1) and (4).

5.4.3 Combining Sequential and Temporal Patterns

We can see that the sequential and temporal patterns mentioned above are all based on our understanding of daily human routines. For example, if asked to predict where a friend (Bob) will go next, we may say ‘to the gym’ if we know that he often visits the gym on Tuesday evenings or after staying at home for the whole afternoon. However, human behaviors are diverse, and we cannot rely on any single set of rules to explain and predict all behaviors. Bayesian inference (Do et al., 2015), random forest (Khoroshevsky & Lerner, 2016), and embedding theorem of non-linear time series analysis (Scellato et al., 2011) were used in previous work to combine patterns to improve prediction accuracy. The method based on random forest in (Khoroshevsky & Lerner, 2016) does not beat the MC1-based method. In (Do et al., 2015), several temporal patterns are combined using Bayesian inference. The results show that the prediction accuracy increases when combining more patterns. However, Do et al. (Do et al., 2015) did not include sequential patterns, and additionally, they adopted a next time bin model (see Section 3) instead of the next-place model used in this chapter.

5.5 Methodology

To answer the four research questions, we select three single pattern-based methods (one temporal pattern and two sequential patterns) and two hybrid methods combining the single patterns. Figure 5.2 shows the five methods to be compared in our evaluation: (1) The visiting time distribution of each place for the given time of day and day of the week (Temporal); (2) MC1; (3) Active LeZi; (4) Joint conditional probability with both ○1 and ○2 (Temporal+MC1); (5) Joint conditional probability with both ○1 and ○3 (Temporal + Active LeZi).

<table>
<thead>
<tr>
<th>First-order Markov Chain (MC1)</th>
<th>Visiting Time Distribution on Each Day of the Week (Temporal)</th>
<th>Active LeZi (ALZ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>○1</td>
<td>○2</td>
<td>○3</td>
</tr>
<tr>
<td>○4</td>
<td>○5</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5.2: The five selected methods in our evaluation.**

Method ○1 represents the only temporal pattern-based method in our evaluation. Given the current time in minutes of a day ($t_a$) and day of the week, we calculate the visiting frequencies of each candidate place on this day of the week after $t_a$ in the current day to get the conditional probabilities ($P_i = P(l_i|c = t_a)$) for POI $l_i$.  

71
We select the place with the highest visiting frequency as the prediction. If any \( P_i \) is zero, we assume that the user stays in the current place for the rest of the given day (e.g., home) and we then use the next day of the week to calculate the conditional probabilities. Method ② and ③ use the first-order Markov chain and Active LeZi, respectively, which have been described in Section 5.4.1.

\[
\arg \max_i P(l_i|c) \tag{5.1}
\]

As shown in Formula (5.1), method ①, ② and ③ choose the POI \( l_i \) that maximizes the conditional probability \( P(l_i|c) \) as the prediction. The \( c \) refers to the temporal or sequential pattern.

Method ④ and ⑤ combine the conditional probability based on the temporal pattern with those based on the two sequential patterns. Under the naïve assumption that the temporal pattern and sequential patterns are independent of each other, we calculate the joint conditional probability using Bayes’ theorem as shown in Equation (5.2).

\[
P(l|c) = \frac{p(c|l)p(l)}{p(c)} = \frac{\prod_i p(c|l_i)p(l)}{\prod_i p(c_i)} = \frac{\prod_i \left( \frac{p(c|l_i)p(l)}{p(l)} \right)}{(p(l))^{N-1}}
\]

\[
= \prod_i \left( \frac{p(c|l_i)}{p(l)} \right)^{N-1} \tag{5.2}
\]

The POI with the maximum joint conditional probability \( P(l_j|c) \) is regarded as the prediction as shown in Formula (5.3), where \( l_j \) refers to the j-th prediction candidate.

\[
\arg \max_j P(l_j|c) \tag{5.3}
\]

### 5.6 Evaluation

This section presents the evaluation results of the selected methods. We begin by describing the datasets, and then introduce the incremental-prediction accuracy and discuss the predictability limit. All the methods were implemented and evaluated in Matlab 2017a.

#### 5.6.1 Datasets

The two datasets used in our evaluation are the *Mobile Data Challenge (MDC)* dataset (Kiukkonen et al., 2010; Laurila et al., 2012) and the *StudentLife* dataset (R. Wang et al., 2014). Both datasets contain daily life mobility data collected by smartphones over a 24-hour period, seven days per week. The data collection duration in the MDC dataset is 18 months and thus much longer than in the StudentLife dataset (10 weeks). To make the MDC dataset comparable to the StudentLife dataset, we exclusively selected the participants’ data from the MDC dataset covering more than 60 days, which resulted in a dataset containing data...
from 144 participants. Moreover, we only kept 60 days of data for each participant for this dataset.

We obtained the sequences of POIs based on these two datasets using the POI clustering method proposed in Chapter 4 (Y. Wang, Sommer, et al., 2018). Each user’s mobility is modeled as one sequence of POIs, as shown in Table 5.1. We use the term *sequences* for short to indicate ‘sequences of POIs’ in the remainder of this section.

Since prediction results strongly depend on the quality of the extracted sequences (Khoroshevsky & Lerner, 2016), we first filter the sequences according to the following criteria: (1) sequences containing less than three or more than 90 unique places are ignored; (2) sequences with a length shorter than 50 or over 3,000 are excluded. A very small number of unique places or a short sequence length indicate that the sequences might miss significant mobility data. On the other hand, a very large number of unique places or long sequence length can indicate that the sequences contain too much noise. We apply our criteria to the resulting 49 sequences from the StudentLife dataset and the 144 ones from the MDC dataset, respectively. Twelve sequences had less than three unique places, while thirteen sequences were longer than 3,000. Finally, 47 sequences from StudentLife dataset and 71 sequences from MDC dataset were kept for evaluation.

![Figure 5.3](image)

**Figure 5.3: Distribution of sequence length and POI number of the prepared data from the StudentLife dataset and MDC dataset.**

Figure 5.3 shows the distribution of sequence lengths and the number of unique places for the selected sequences from both datasets. The majority of sequences from the StudentLife dataset have 100-300 data points and the sequences include 3-17 POIs. The MDC dataset contains many longer sequences compared to the StudentLife dataset. However, the range of the POI number (3-18 in most cases) in the sequences from the MDC dataset is similar with that from the StudentLife dataset, which is reasonable.
5.6 Evaluation

5.6.2 Incremental-Prediction Accuracy and Times of Winning

To simulate the real-time performance of the selected methods, we used Incremental-Prediction Accuracy (IPA) as an evaluation metric. From the beginning of each sequence, we predict the next place based on the previous information. The models in each method will be updated once a new data point is available. The IPA is calculated as in Formula 5.4, where \( n_{\text{correct}}(t) \) is the number of correct predictions up to the current moment \( t \) and \( n_{\text{total}}(t) \) is the number of total predictions up to \( t \).

\[
P(t) = \frac{n_{\text{correct}}(t)}{n_{\text{total}}(t)}
\]  
(5.4)

In some related work, conventional validation (i.e. partitioning the data set into two sets, e.g. 70% for training and 30% for testing) is used (Burbey & Martin, 2012; S. Lee, Lim, Park, & Kim, 2016). However, this method does not allow analyzing the performance over time from the beginning to the end of data collection. Cross-validation is also not suitable for sequential prediction, especially when using sequential patterns (e.g., MC1 and Active LeZi), because the sequential information of human behavior will be broken when the sequence is segmented or sampled.

Along with IPA, we introduce another metric called Times of Winning (ToW) as shown in Formula (5.5). \( P_m(t) \) means the IPA of a given method, while \( \{P_\emptyset(t)\} \) indicates the IPA of each used method. Therefore, ToW refers to the frequency of the cases where a method beats others over the given sequence.

\[
\text{ToW}_m = \text{number of } (P_m(t) > \{P_\emptyset(t)\})
\]  
(5.5)

5.6.3 Predictability Limit

To compare the selected methods, we are not only interested in the relative performance of the methods. Instead, we also examine their distance to the predictability limit. The upper limit of human mobility predictability represents the average probability of correctly predict the next POI that can be best achieved, given a POI sequence. Depending on the entropy rate of a given sequence, the predictability limit can be estimated. Song et al. (Song et al., 2010) provided an approach for estimating the theoretical predictability limit of sequential data using LZ-based entropy rate estimation (Kontoyiannis, Algoet, Suhov, & Wyner, 1998) and Fano’s inequality (Fano, 1949). The predictability limit provides us with a theoretical benchmark for estimating how good a prediction method works.

5.6.4 Accuracy@K

To analyze the average accuracy of each single pattern-based method, we also use the Accuracy@K metric which indicates the rate that the top K prediction candidates contain the correct prediction. This metric has been well used in
location prediction and location-based recommendation (Xie et al., 2016; Z. Zhang et al., 2017). In our evaluation, we use K as 3.

5.6.5 Results

Before showing the results of the average prediction for all users in each dataset, we first show the IPA of each method for two users’ data from each dataset in Figure 5.4. The figure shows us that the IPAs of all methods fluctuate during a short period of time in the beginning. After this fluctuation, most methods arrive at a relatively stable IPA level with a slight increase as more information is obtained. We can also observe that there are some common trends of short and rapid decreases in the middle of the process, which indicates a change of the users’ daily routines. Although the predictability limit (the horizontal black dotted lines in Figure 5.4) is only subject to the sequential pattern-based prediction, it also cannot be reached by other methods. For both users, the temporal pattern-based method (green line in Figure 5.4) performs the worst, while the best-performing method changes over time.

![Figure 5.4: The results of incremental-prediction accuracy (IPA) for user #2 from StudentLife dataset (left) and user #8 from MDC dataset (right).](image)

The average prediction accuracies of each method for both datasets are shown in Figure 5.5. All the reported numbers represent the median values in the remainder of this section. For the StudentLife dataset, the results of method ⊕ (MC1) and method ⊗ (ALZ) are comparable (0.41), while being significantly better than method ⊖ (0.22), ⊗ (0.30), and ⊖ (0.26). When comparing only the two sequential pattern-based methods, method ⊕ (MC1) performs slightly better than method ⊗ (ALZ) in the perspectives of the interquartile range and whiskers of the box plots. Combining temporal and sequential patterns cannot improve the average prediction accuracy, which is out of our expectation. For the MDC dataset, method ⊕ (0.71) outperforms method ⊗ (0.62), which is different than the result for the StudentLife dataset. The temporal pattern-based methods still performed the worst, and the average prediction accuracy of the methods combining temporal and sequential patterns was worse than the methods using single sequential pattern. All selected methods were far from reaching the predictability limits (PL) for both datasets.
5.6 Evaluation

Given our results so far, research question Q1 can be answered by stating that the selected sequential pattern-based methods outperform the selected temporal pattern-based method. From the view of average accuracy, the hybrid methods do not perform better than their counterparts that only used sequential patterns. In Figure 5.4, we already saw that the winner (i.e., the method with the highest IPA in real-time) changes over time.

From the perspective of ToW, method $\odot$ (MC1+Temporal) is best for the StudentLife dataset, while method $\odot$ (ALZ) outperforms the others for the MDC dataset. Therefore, to answer research question Q2, we argue that combining sequential with temporal patterns does not necessarily improve the average prediction performance but can outperform other methods only for some time periods for specific users.

In addition, research question Q3 can be answered by our observation that there is no consistently winning method among the evaluated methods. When considering ToW, method $\odot$ (Temporal) and method $\odot$ (Temporal + ALZ) cannot beat the other methods for most time intervals for both datasets.

In terms of the Accuracy@3 results, we can see that MC1@3 achieves the highest accuracy (0.77) in the StudentLife dataset, while ALZ@3 (0.93) performs slightly better than MC1@3 (0.92) for the MDC dataset, as shown in Figure 5.5. If we
compare each @3 accuracy of MC1 and Active LeZi with their top-1 prediction accuracy respectively, MC1@3 improves more than ALZ@3 for both datasets.

5.7 The Visualization Tool UP³V

Only using the IPAs of the tested methods at the average-level of individuals is not enough to help developers and system designers answer the question: Which method(s) shall we choose in real-world personalized applications, e.g. JITAI’s? This question is precisely the aim of research questions Q4. To help experts analyze the performance of each method at the level of the individual, as well as at the POI level, we developed a visualization tool, which we termed the User’s Pattern and POI Prediction Visualizer (UP³V). It interactively visualizes the results for each user’s data. The target users of this tool are the experts who analyze human mobility patterns. This proof-of-concept prototype is implemented based on Shiny7 in an RStudio environment. Leaflet8 and graphics9 were used for the map and charts.

Figure 5.6 shows UP³V containing the map with spatial information (left side) and the information panel (right side), which contains the statistical visualization showing temporal, as well as sequential results. Once a POI – represented as a cluster of dots in a specific color on the map – is selected, all the transitions from other places to the selected POI will be shown on the map. The information panel contains charts presenting prediction results at the level of each user (as shown in Figure 5.4) and each POI (see chart ○ in Figure 5.6). To support temporal patterns analysis, two charts are generated (chart ○ and ○) presenting the arrival time distribution per weekday of the selected POI and the stay periods for the selected POI in the visualization tool. Chart ○ and ○ can support experts to discover temporal patterns. For example, from chart ○ we can deduce that the selected POI should be the home of the user, because the user typically remained in this place during the night (the yellow cluster on the map). Further, from chart ○, we can see that the arrival time at this place was widely distributed for different days of the week, e.g. the most frequent arrival time on Wednesdays was close to midnight.

The panel also has a data selector for switching the dataset and the user ID. When the dataset or user ID is selected, the map and the figures update accordingly. When any POI is clicked, the corresponding marker will appear on the map and the result of incremental-prediction accuracy in the POI level will show up as in chart ○ in Figure 5.6, which helps us answer research question Q4. Using UP³V, we found that the accuracies of different method vary depending on the current POI of a specific user. Furthermore, the result shows that the performance of different methods changes and the winner varies over time for a selected POI.

---

7 https://shiny.rstudio.com/
8 https://rstudio.github.io/leaflet/
9 https://www.r-graph-gallery.com/
5.8 Discussion

Using this visualization tool, an expert can gain a deeper understanding of each user’s data in terms of the information available on spatial, temporal, and sequential patterns for each POI. We find that no method can consistently outperform other methods for all places visited by a certain user. For sequences of most users, the winner changes over time. Besides, given a time point over the sequence of a user, we find that the winning method for a specific POI (see chart ○ in Figure 5.6) can be different from the winning method for all POIs on average for the user (see Figure 5.4).

![Figure 5.6: User’s Pattern and POI Prediction Visualizer (UP3V). The map (left side) displays the places/spatial information and an information window (right side) shows the results per selected user. The information panel shows for the selected POI: (1) the prediction accuracy, (2) the arrival time distribution per weekday, and (3) the stay periods.](image)

5.8 Discussion

Based on the initially mentioned research questions, the basic idea of this work was to test if integrating temporal and sequential patterns using joint conditional probability is able to improve the prediction accuracy. Our results did not show performance improvement on average. However, the hybrid methods (○ and ○) can beat single sequential pattern-based methods (○ and ○) for some periods of the data sequence in terms of incremental-prediction accuracy. When considering the temporal pattern-based method we adopted, the performance was quite low.

Another interesting finding was that the more sophisticated sequential patterns (e.g., Active LeZi) were not always more effective than simple ones (e.g., MC1). In addition, the performance of the algorithms highly depends on the corresponding dataset. Based on two datasets (128 users in total), our methods focused on three single patterns (two sequential patterns and one temporal pattern). Larger datasets may change the results of prediction at the average
level, but the findings at the level of the individual and at the POI level using IPA and ToW will still stand, which will be relevant for future approaches. In the following, the implications based on the results of our evaluation and the discoveries made with UP3V are summarized:

(1) We suggest that methods based on different patterns are used in parallel for next-place prediction because individuals’ mobility data is quite diverse over time. JITAI designers can always choose the winning method based on incremental-prediction accuracy to take full advantage of different methods. If several POIs are examined, using the winning method over time for the corresponding POIs is recommended instead of using the winning method for all POIs on average.

(2) When using top 3 predictions (Accuracy@K), the prediction accuracy can be much higher than for top 1 prediction when using MC1 or Active LeZi.

(3) From our observation using UP3V, we found that the stay periods of some POIs also have embedded patterns, which can be potentially used for improving the prediction accuracy.

5.9 Conclusion

In this chapter, we addressed the problem of next-place prediction to provide insights on related mobile health applications. Based on individuals’ sequential and temporal patterns in the mobility data collected by smartphones, we posed four research questions regarding the effectiveness of three single pattern-based methods (first-order Markov chain, Active LeZi, and an arrival time distribution-based temporal pattern) and two methods using joint conditional probability to combine temporal and sequential patterns. Our primary contributions include:

(1) we evaluated the performance of combining sequential and temporal patterns when compared to the use of single patterns, (2) we used the predictability limit to compare the prediction performance of the selected methods, (3) we investigated the performance of next-place prediction methods at the level of the individual and at the level of POIs (Places of Interest) using incremental prediction accuracy and times of winning.

We evaluate the five selected methods at both of these levels for two publicly available datasets. Using incremental prediction accuracy, we find that each of the selected methods performs differently among users and POIs over time. On average, the sequential pattern-based methods perform better than the temporal pattern-based methods. Using joint conditional probability to integrate different patterns is effective to some extent but does not outperform single pattern-based methods on average. In addition, the top 3 predictions (Accuracy@K) of sequential pattern based-methods show high accuracy. Our results suggest that using multiple methods in practice, and that relying on the best performing method for each individual and even for each POI depending on incremental prediction accuracy, can lead to an optimal outcome.
In addition, the visualization tool UP^3V was presented, which is designed to analyze the prediction results per user and per POI. In future work, we plan to extend and evaluate our visualization tool with domain experts. Additionally, we will compare more temporal and sequential patterns and explore how to effectively combine these patterns.

User-mobility prediction could enable the study of opportune moments for reminding in sedentary behavior change interventions. However, due to the large predictability diversity of individuals’ mobility based on the smartphone-logged data, we did not apply next-place prediction in our interventions studies.
6 Supporting Sedentary Behavior Change by Visualizing Personal Mobility Patterns and Action Planning on Smartphone

“凡事预则立，不预则废。

Preparedness ensures success, and unpreparedness spells failure.

- 《礼记》
  Book of Rites
  (A Confucian classic)


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Laura Koenig helped in conducting the study, discussing the data analysis, and reviewing the paper. Harald Reiterer supervised the work and reviewed this paper.
6.1 Abstract

Prolonged sedentary behavior is related to a number of risk factors for chronic diseases. Given the high prevalence of sedentary behavior in daily life, light-weight solutions for behavior change are needed to avoid detrimental health effects. Based on behavioral theories, the mobile app SedVis was developed. The app provides personal mobility pattern visualization and action planning. We aimed to study the effect of the mobility visualization on users’ action planning and change in sedentary behavior, as well as engagement with the visualization and user experience of the app. We conducted a three-week pilot study with 16 participants who had the motivation to reduce their sedentary behavior. A mixed study design was adopted: after the one-week baseline period with no access to the functions in the app, only the intervention group (N=8) was given the access to the visualizations, while both the control group (N=8) and the intervention group were asked to make action plans every day during the two-week intervention period. We analyze the data using both traditional null-hypothesis significance testing and Bayesian statistics. The results showed that the visualizations in SedVis had no statistically significant effect on the participants’ action planning. The intervention involving the visualizations and action planning in SedVis had a positive effect on reducing participants’ sedentary hours with weak evidence, while the control condition did not decrease sedentary time. We also found that the more frequently the users checked the app, the more they reduced the sedentary behavior. The visualizations in the app also led to higher user-perceived novelty. No participants complained about the interruption, while some participants commented that making action plans every day was boring. Using a smartphone app to collect mobility data and provide feedback in real time using visualizations is a promising method to induce changes in sedentary behavior and might be more effective than action planning alone.

6.2 Introduction

6.2.1 Background

The high prevalence of opportunities to be sedentary in our daily life leads to high habit strength of sedentary behavior, which makes it difficult to change in the long term (D. W. Dunstan et al., 2010). Interventions are therefore needed to support individuals to reduce their sedentary time. In their review, Chu et al. (Chu et al., 2016) divided intervention strategies of reducing sedentary behavior to three categories: (1) educational/behavioral (e.g., goal setting, action planning, and self-monitoring); (2) environmental changes (e.g., sit-stand workstation and treadmill desk); (3) multi-component (e.g., sit-stand workstation plus goal-setting). The environmental and multi-component interventions might require policy support and additional facilities, which might hinder their immediate
application on a larger scale. Therefore, low-cost and light-weight solutions are needed.

Mobile devices including smartphones and wearables (e.g., smartwatches and fitness wristbands) might be a solution to this problem. Firstly, the prevalence of both smartphone and wearable device ownership is increasing globally (Statista, 2019). Secondly, these devices allow monitoring physical activity and sedentary behavior. When fed back to the user, the data might help them to generate meaningful insights about their activity patterns and subsequently induce behavior change. While self-monitoring and feedback are behavior change techniques (Abraham & Michie, 2008; Michie et al., 2013) commonly used in digital interventions for physical activity and sedentary behavior (Middelweerd et al., 2014; Schoeppe et al., 2017), the present study aimed to extend this approach by using an interactive visualization of sedentary behavior data to specifically support daily action planning, thereby facilitating a reduction of sedentary time in daily life.

6.2.2 Action Planning for Sedentary Behavior Change

An action plan combines specific situation parameters (“when” and “where”) and a sequence of actions (“how”) for a target behavior (Schwarzer, 2008). In this vein, it is suggested that a behavior will be triggered automatically when encountering the specific situation (Gollwitzer, 1999). Action planning might therefore bridge the gap between intention and behavior [10]. Indeed, several meta-analyses have shown that action planning is positively related to goal attainment and health behavior change (e.g., (Gollwitzer & Sheeran, 2006; Michie et al., 2009; Olander et al., 2013)) and thus might be an effective behavior change technique.

To date, there are only a few studies that included action planning in digital interventions targeting sedentary behavior (Stephenson, McDonough, Murphy, Nugent, & Mair, 2017). In our systematic review of digital technologies supporting health behavior change (see details in Chapter 9) (Y. Wang, Fadhil, & Reiterer, 2019), only two out of 45 reviewed studies involved action planning related to sedentary behavior change. Based on step counts at baseline, Aittasalo et al. (Aittasalo et al., 2017) offered the participants visual feedback to facilitate their action planning, while De Cocker et al. (De Cocker, De Bourdeaudhuij, Cardon, & Vandelanotte, 2015) used several motivational questions to stimulate the participants to make action plans. In both studies, sedentary behavior was successfully reduced. However, both used action planning as one of several behavior change techniques, and it is therefore unclear whether the change can be solely attributed to action planning. Maher and Conroy (Maher & Conroy, 2015), on the other hand, specifically tested the main effect of action planning on reducing sedentary behavior and found that daily action planning did not induce sedentary behavior change. This study, however, has limitations. First, sedentary
behavior was only assessed subjectively, which might not correspond to objectively measured behavior (Hagstromer, Ainsworth, Oja, & Sjostrom, 2010). Second, the quality of the action plans was not evaluated. This, however, might have provided important insights on why the intervention was not successful.

The quality of an action plan can be evaluated based on plan characteristics such as specificity of the situational parameters, plan instrumentality, i.e., the degree to which a plan is helpful to achieve a desired outcome, and viability, i.e., how realistic an action plan is. Fleig et al. (Fleig et al., 2017) showed that specificity of when to perform a behavior and instrumentality of the action plan were related to an increased likelihood of plan enactment. Quality of action plans therefore might be an important variable to consider when evaluating interventions. While none of the aforementioned studies on sedentary behavior change investigated the quality of action plans, the present study aimed at testing the effect of action planning on sedentary behavior change quantitatively and additionally included a qualitative analysis of the action plans to determine their specificity, instrumentality, and viability. Building upon the idea of Aittasalo et al. (Aittasalo et al., 2017) that visualizations of sedentary behavior data might facilitate action planning, we developed a novel tool to support action planning for reducing sedentary behavior using interactive visualization.

6.2.3 Visualizations of Mobility Patterns

We developed a mobile app – SedVis – aiming to reduce users’ sedentary time through visualizing the user’s mobility patterns (i.e., the time and the locations/routes of both sedentary and active sessions) and action planning. SedVis automatically tracks and classifies users’ activity (e.g., walking, biking, and in a vehicle), steps count, and locations, and in this vein determines locations and time windows in which users are sedentary. The visualization elements thus correspond to the aforementioned action planning factors - when, where, and how (i.e., the planned activity). The visualization can then be used as a basis for formulating action plans. To the best of our knowledge, SedVis is the first app targeting visualizations and action planning on mobile devices for sedentary behavior change.

6.2.4 Study Objectives

In this chapter, we report the results of our three-week pilot study of SedVis (N=16). First, we aimed to examine the effect of SedVis on users’ action planning for their sedentary behavior change (RQ1). Specifically, we were interested in whether using the visualization improved specificity, instrumentality, and viability of formulated action plans. Second, we tested whether the intervention involving visualizations and action planning is effective in reducing sedentary behavior compared to action planning without visualizations (RQ2). Third, because the designed visualizations could also serve as a self-monitoring tool, we aimed to investigate users’ engagement with the visualizations in SedVis and its
impact on users’ sedentary behavior change (RQ3). Four, we investigated user acceptance and experience of SedVis as a light-weight intervention tool for the daily use of the sedentary population (RQ4).

6.3 Methods

6.3.1 Study Design

We deployed a mixed study design (see Figure 6.1). Participants were assigned to two groups (between-subjects factor): Group A, for which the visualization functions were enabled; Group B, for which the visualization functions were disabled. Participants were assigned to the groups according to the enrollment time (i.e., every odd number was assigned to Group A, while every even number was assigned to Group B). This strategy enables fast study-deployment for each participant while keeping the balance of sample size in both groups. Our study included three interviews (i.e., the entry interview before starting data collection, the baseline interview after week 1, and the exit interview after week 3) on day 1, day 9, and day 25 for each participant. The data collected on these three days in the app were excluded in data analysis.

**Figure 6.1: The mixed design of our three-week study.**

During the entry interview, the participants were informed about the purpose of the study, signed the consent form, and filled out questionnaires on demographics and psychosocial variables related to sedentary behavior. An experimenter then installed SedVis on their smartphones.

The baseline interview took place the day after the baseline week. Participants filled out another questionnaire on psychosocial variables before watching an educational video about the risks of prolonged sedentary behavior. Subsequently, the experimenter showed them a flyer to explain a behavior change theory (Schwarzer, 2016) and emphasized the importance of action planning. The participants were asked to make at least one plan per day to reduce their sedentary behavior for the following two weeks. Finally, the experimenter introduced the functions in the app, depending on which group participants of the session were assigned to. For Group A, all the functions were activated, including daily visualization, multi-day visualization, and action planning. For Group B, only the action planning function was enabled (see more details below).

---

10 https://youtu.be/wUEI8KrMz14
For both groups, participants were asked to set a daily reminder within the app to make action plans.

After two more weeks, the participants returned to the lab for the exit interview, when they completed another questionnaire on psychosocial variables, transferred the data stored on their smartphones, and took part in a short semi-structured interview. Each participant received 20€ after completing the study.

The ethics committee of the University of Konstanz approved the study protocol. For privacy reasons, we only collected the data related to the study. To ensure transparency of data collection, we only recorded the data on the participants’ smartphones and showed them the data details when they transferred the data.

6.3.2 Participants

We recruited participants through the university mailing lists, the authors’ social media profiles, and posters in the university. Sixteen participants met the inclusion criteria: (1) having the intention to change their sedentary behavior; (2) having no injuries that precluded them from being physically active; (3) being able to speak English fluently; (4) owning a smartphone with Android 6.0 and above; (5) not using a standing desk; (6) having no travel plans during the study period. The fifth criterion was used to filter out people who already started to change their sedentary behavior. The other criteria were used to control the motivation and objective ability for using our app, communicating with the experimenters, and changing sedentary behavior.

All participants were students (9 PhD students; 9 female) at the university. Group A comprised of five females and three males. Their mean age was 26.6 (standard deviation = 3.8). Group B comprised four females and four males. Their mean age was 27.0 (standard deviation = 4.0). Among the 16 participants, one was overweight (i.e., BMI > 25), one was underweight (i.e., BMI < 18.5), and the remaining had a normal weight (mean = 22.0, standard deviation = 2.8).

6.3.3 SedVis App

6.3.3.1 Data Collection

SedVis was developed for Android smartphones. It collected the data of physical activities (via Google Activity Recognition API), geolocation (via Google Maps API), steps (via Google Fit API), screen states (turned on or off), users’ interaction within the app, users’ action plans, and time stamps. High battery consumption (e.g., through constant geo-location tracking) or large disk-space requirement might lead to users’ abandonment of the app. To avoid dropout, geolocation was only updated when movements were detected based on activity recognition and steps counting every five seconds. Besides, a new data point was only recorded when a change of the activity state was detected (e.g., the steps increase or the physical activity changes). This strategy minimizes energy consumption and
data storage without losing information on users’ mobility (Y. Wang, Duan, et al., 2016).

To improve power consumption, Google imposes limitations on background services since Android 8.0. Some OEM versions (e.g., MIUI and EMUI) of Android additionally introduced limitations on background services to optimize the battery life. These limitations presented a side effect of our state-based data collection strategy: The historical data did not allow to determine if the background service kept running. Therefore, we added a timer to the background service to log a timestamp to the local database every 20 minutes. To improve the data collection quality, data collection service was bound to a notification showing the latest update time, steps, and activity in the notification bar (as shown in Figure 6.2). A system clock was used to monitor if the background service is running and, if necessary, to initiate a restart. Users could also manually re-start the data collection service if the notification disappeared.

Figure 6.2: The always-on notification of SedVis on a user’s smartphone.

6.3.3.2 Dashboard

From the dashboard of SedVis, participants could access all the functionalities of the app, as shown in Figure 6.3. In the settings tab, the experimenter could enable intervention functions. Passwords were used to restrict the users’ access to these functions during our study.

https://developer.android.com/about/versions/oreo/background
https://en.miui.com/
https://consumer.huawei.com/en/emui/
6.3 Methods

![SedVis Dashboard](image)

**Figure 6.3:** The dashboard of SedVis.

6.3.3.3 Mobility Pattern Detection and Visualization

Mobility patterns refer to the when, where, and how the user moves. In SedVis, this involved tracking of users’ moving trajectory and sedentary place detection. The trajectories showed the routes the user had taken and related information on step counts and time windows. The app detected the users’ physical activity every five seconds, which enabled a high temporal resolution for trajectory tracking. Modern smartphones use high-precision and low-power movement sensors, which makes the physical activity recognition and steps tracking both accurate and efficient. Google Play services provide fused location tracking by using GPS, Wi-Fi, and cellular signals to allows for precise positioning even in some indoor environments\(^\text{14}\).

Custom programmed sedentary place detection was used to detect the participants’ sedentary places based on the users’ geolocation data. Many office workers spend the day in a limited number of locations (e.g., home, office, lab) where they spend much time sitting. Existing services, like the Places SDK for Android\(^\text{15}\), only provide public places (e.g., the university), which could not enable personalized places detection in other places such as at home. Therefore, we used the spatio-temporal data clustering algorithm proposed in Chapter 4 (Y. Wang, Sommer, et al., 2018) to detect the places based on each user’s data. These detected sedentary places provide users intuitive hint on where to reduce their sedentary behavior.

\(^{14}\) It depends on the strength of the indoor Wi-Fi and cellular signals.

\(^{15}\) [https://developers.google.com/places/android-sdk/intro](https://developers.google.com/places/android-sdk/intro)
Within SedVis, users could access two visualizations of data on their sedentary and active hours. We used 250 steps/hour as the threshold for sedentary hours as the Fitbit mobile application. In the daily visualization, the tracked trajectories and the detected sedentary places were shown on a map, and the corresponding temporal information was shown using a bar chart (as shown in Figure 6.4) for a single day. In addition to access via the dashboard, single-day visualization could be accessed by clicking the always-on notification (see Figure 6.2). In the daily visualization, users could see (1) the active hour(s) and the corresponding routes on the map once clicking on a blue bar, and (2) the sedentary hour(s) and the corresponding locations once clicking on an orange bar. Likewise, clicking the sedentary location on the map highlighted the corresponding sedentary hours in the bar chart. While the bar chart illustrated temporal patterns, the map demonstrated spatial patterns. Participants could switch between days by tapping on the arrows at the bottom of the screen.

![Figure 6.4: The mobility patterns in the daily visualization mode.](image)

In the multi-day visualization, sedentary places were determined based on aggregated data from the user-selected days. Differing from the daily visualization, the bar chart in the multi-day visualization showed the frequencies the user was sedentary in each hour during the selected days for all the places or one selected place (see Figure 6.5).

---

16 [https://www.fitbit.com/app](https://www.fitbit.com/app)
6.3 Methods

6.3.3.4 Action Planning

We use a list view to show the action plans the user has made. When adding an action plan, the user is expected to specify the “when”, “where”, and “how” elements (see Figure 6.6). The user can enter the action planning view from the dashboard or through the shortcuts in the visualization views (see the second button in the left-up corner of Figure 6.4 and Figure 6.5). The action plans are shown chronologically and cannot be deleted.

Figure 6.5: The multi-day visualization.

Figure 6.6: The action planning function in the app.
6.3.4 Measures

6.3.4.1 Sedentary behavior

The sedentary hours were calculated based on the steps counts assessed by the SedVis app. Sedentary behavior was quantified per hour. An hour was labeled as sedentary if less than 250 steps were recorded. It should be noted that the sedentary hours included the participants’ sleeping time. We assumed that the participants’ sleeping time did not change over the three-week study, which was confirmed by the participants in the exit interview.

6.3.4.2 Number of action plans

The total number of action plans formulated during the two-week intervention phase was counted automatically by the SedVis app. Since we allowed the participants to repeat the plans of previous days, we also calculated the number of unique action plans.

6.3.4.3 Quality of action plans

To evaluate the quality of the action plans, we coded the specificity of the When, Where, and How of the plans. The rating criteria for three levels of specificity (i.e., vague, medium specific, and highly specific) were adapted from Fleig et al. (Fleig et al., 2017) (see Table 6.1 for coding criteria). We also asked the participants to evaluate the viability (how realistic) and instrumentality (how useful) of their action plans based on the plan characteristics used by Fleig et al. (Fleig et al., 2017). For viability, we asked participants to rate each action plan on a scale from 1 (not realistic at all) to 4 (very realistic); for instrumentality, we asked participants to rate each action plan on a scale from 1 (not helpful at all) to 4 (very helpful).

Table 6.1: Coding criteria for specificity.

<table>
<thead>
<tr>
<th>“When”</th>
<th>Vague (=1)</th>
<th>Medium Specific (=2)</th>
<th>Highly Specific (=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Today”</td>
<td>Empty; “Now”; “Anytime”; “Today”</td>
<td>“Every Hour”; “After Lunch”</td>
<td>Time point (e.g., “13:00”)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“Where”</th>
<th>Empty; “Out”</th>
<th>Large area (e.g., “City”, “University”)</th>
<th>Places (e.g., “Post”, “Lab”, “Office”, “Home”, “Library”)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>“How”</th>
<th>Empty;</th>
<th>“Going to the park”</th>
<th>Activity (e.g., “Walk”, “Yoga”, “Cycle”, “Pushups”, “Stretch”, “Stand up”)</th>
</tr>
</thead>
</table>

6.3.4.4 Engagement with the app

Participants’ interaction with the app was quantified by recording all operations in the app during the study, including how often the participants checked the
visualizations. In addition, we logged the time stamps of when participants made action plans, which we then used as the basis for discussing the users’ experience with the app during the exit interview.

6.3.4.5 Intention to change sedentary behavior

The participants’ intention was measured using a scale from 1 (I do not plan to reduce my sedentary behavior at all true) to 4 (I do exactly plan to reduce my sedentary behavior) following the example in HAPA (Schwarzer, 2008). We used intention as a control measure as we required that the participants joined the study due to their motivation of reducing their sedentary behavior instead of other factors.

6.3.4.6 User experience

Using the user experience questionnaire (UEQ) (Schrepp, Hinderks, & Thomaschewski, 2014), the user experience of the app was quantified at the exit interview. In addition, we used open-ended questions to explore the participants’ attitudes to the app and the study, as well as their desired features missing in the app.

6.3.5 Statistical Analysis

Data were analyzed using both traditional null-hypothesis significance testing (i.e., t-tests, ANOVAs, and Pearson’s correlation coefficient) as well as Bayes factors. The conventional null-hypothesis significance tests provide little information when the result is not statistically significant – only the alternative hypothesis is tested (Dienes, 2014). Non-significant results might support a null hypothesis over the alternative, or the data are just insensitive. By contrast, Bayes factors (Kass & Raftery, 1995) compare the extent to what the samples support two hypotheses (e.g., equal or different). Besides, Bayesian methods also allow more principled conclusions from small-n studies of novel techniques in the field of human-computer interaction (Kay, Nelson, & Hekler, 2016). Therefore, we use the Bayes factor (BF) in addition to the p-value (Greenland et al., 2016) and Cohen’s d (C. O. Fritz, Morris, Richler, & Fritz, 2012) to report and interpret the results. We use JASP\textsuperscript{17} (Version 0.9.2) for data analysis due to its ability of both the conventional null-hypothesis significance test and the corresponding Bayesian analysis.

The Bayes factor is a ratio of the likelihood probabilities. $P(H_0 \mid data)$ is the probability of the null hypothesis ($H_0$) given the data, while $P(H_1 \mid data)$ is the probability of the alternative hypothesis ($H_1$) given the data. The definition of the Bayes factor is shown in Formula 6.1 as below.

\textsuperscript{17} https://jasp-stats.org/
Chapter 6: Supporting Sedentary Behavior Change by Visualizing Personal Mobility Patterns and Action Planning on Smartphone

\[ BF_{01} = \frac{P(H_0 \mid data)}{P(H_1 \mid data)} \quad \text{or} \quad BF_{10} = \frac{P(H_1 \mid data)}{P(H_0 \mid data)} \]  (6.1)

The Bayes factor indicates which hypothesis is more supported by the data. Figure 6.7 shows the Bayes factor classification and the interpretation we adapted from (Doorn et al., 2019). The default priors of the alternative hypothesis and the calculation methods for different study design can be found in the work of Rouder and colleagues (J. N. Rouder, Morey, Speckman, & Province, 2012; J. Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Figure 6.7: A graphical representation of a Bayes factor classification and the interpretation, adapted from (Doorn et al., 2019).

We use the default Cauchy distribution \( r = 1/\sqrt{2} \) as the prior when estimating the effect size. Following the JASP guidelines (Doorn et al., 2019), we also report the median (M) and the 95% credible interval (CI) of the effect size. For correlation analysis, we use Bayesian Pearson correlation test with default prior suggested by Rouder and Morey (J. N. Rouder & Morey, 2012). Depending on the context, we report the one-side Bayes factor (BF_{0} or BF_{+0}) or the two-side Bayes factor (BF_{01}).

6.4 Results

6.4.1 Data Collection

All participants completed the study. First, data quality was checked based on the actual running duration to ensure that all participants had access to the app as planned. The missing duration may be caused by smartphone being switched off or the background service being shut down for battery optimization. Only the data from one participant in Group A (A8) showed a relatively low coverage (65.61% of the study duration); for all other, the coverage was 93.88% (SD 5.44%). After checking the data of participant A8, we found they had the habit of shutting down the phone during the night. Therefore, the missing data does not limit conclusions about the mobility of the participant and the participant’s data was analyzed as planned.

6.4.2 Participants’ Intention (Control Measure)

The participants’ intention of reducing sedentary behavior was generally high (mean = 3.20, SD = 0.59) in both groups. No significant difference was found between groups at each appointment according to t-tests (appointment 1: \( t_{14} = 0.00, P = 1, \) Cohen \( d = 0.00 \); appointment 2: \( t_{14} = -0.37, P = .72, \) Cohen \( d = -0.19; \)
6.4 Results

appointment 3: \( t_{14} = 0.00, P = 1, \text{Cohen } d = 0.00 \). The Bayes Factors showed evidence preferring \( H_0 \) (appointment 1: \( BF_{01} = 2.34 \); appointment 2: \( BF_{01} = 2.23 \); appointment 3: \( BF_{01} = 2.34 \)). The result indicated that the participants in both groups had similarly strong intentions.

6.4.3 RQ1: Effect of Visualization on Participants’ Action Planning

First, we aimed to investigate the effect of the visualizations on participants’ action planning. We evaluated both the quantity and quality of the action plans. Although the mean of the number of action plans was higher in Group A (mean = 8.88, SD = 5.69) than in Group B (mean = 7.75, SD = 6.76), the result of t-test shows no significant group difference (\( t_{14} = 0.36, P = .72, \text{Cohen } d = 0.18 \)). The Bayes factor (\( BF_{01} = 2.24, M = 0.11, \text{CI} = [-0.66, 0.92] \)) showed weak evidence towards no difference (\( H_0 \)). It was the same case for the number of unique action plans (\( P = .40, \text{Cohen } d = 0.44, BF_{01} = 1.81, M = 0.28, \text{CI} = [-0.50, 1.14] \)); the mean in Group A was 3.75 (SD = 2.19); the mean in Group B was 2.75 (SD = 2.38).

The quality of the action plans showed mixed results, as shown in Table 6.2. The means of the perceived viability and instrumentality are slightly higher in Group B than in Group A, however, again t-tests showed no significant differences. The Bayes factor showed weak evidence towards difference (\( H_1 \)) of the perceived viability, while no difference (\( H_0 \)) of the perceived instrumentality. The means of the specificity (“When” and “Where”) were higher in Group A than in Group B, although again not statistically significant, while weak evidence supported no difference (\( H_0 \)) according to the Bayes factor. The means of the specificity of the response activity (“How”) were both very high in the two groups because most of the users simply specified the activity as walking.

Table 6.2: The measurements of the quality of the action plans.

<table>
<thead>
<tr>
<th></th>
<th>Group A</th>
<th>Group B</th>
<th>( t_{14} )</th>
<th>( P )</th>
<th>Cohen ( d )</th>
<th>( BF_{01} )</th>
<th>M</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Viability</td>
<td>3.28 (0.68)</td>
<td>3.81 (0.37)</td>
<td>-1.96</td>
<td>.071</td>
<td>-0.98</td>
<td>0.71</td>
<td>-0.67</td>
<td>[-1.72, 0.17]</td>
</tr>
<tr>
<td>Perceived Instrumentality</td>
<td>3.10 (0.55)</td>
<td>3.22 (0.73)</td>
<td>-0.52</td>
<td>.608</td>
<td>-0.26</td>
<td>2.13</td>
<td>-0.15</td>
<td>[-1.00, 0.59]</td>
</tr>
<tr>
<td>Specificity (When)</td>
<td>2.55 (0.70)</td>
<td>1.88 (1.00)</td>
<td>1.59</td>
<td>.14</td>
<td>0.79</td>
<td>1.05</td>
<td>0.51</td>
<td>[-0.30, 1.50]</td>
</tr>
<tr>
<td>Specificity (Where)</td>
<td>2.21 (0.82)</td>
<td>2.10 (0.91)</td>
<td>0.35</td>
<td>.73</td>
<td>0.17</td>
<td>2.24</td>
<td>0.10</td>
<td>[-0.68, 0.94]</td>
</tr>
<tr>
<td>Specificity (How)</td>
<td>2.99 (0.04)</td>
<td>3.00 (0.00)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. For Specificity (How), no results are reported for the t-test and BF because standard deviation in Group B was 0.
Therefore, regarding RQ1, no statistically significant effects of the visualizations in SedVis on the participants’ action planning were found. The Bayes factors indicated weak evidence towards no difference between the two groups except the perceived viability.

Regarding the specificity (“When”) of action plans, we observed some unexpected patterns especially in Group B. Two participants (B3 and B6) in Group B always entered the current time when they made the plan. They explained in the exit interview that each of their plans was actually what they were about to do at the moment when logged the plan. Participant B6 further commented she found it difficult to make action plans for the future because she was not sure about her behavior patterns. Besides, another three participants always used vague cues to specify the “When”: Participant B1 used “today”; Participant B4 used “anytime”; Participant A5 used “today.”

### 6.4.4 RQ2: Changes in Participants’ Sedentary Behavior

A repeated measures ANOVA showed no significant effect of time ($F_{1,14} = 0.14, P = .72$), group ($F_{1,14} = 0.17, P = .68$), or the interaction between time and group ($F_{1,14} = 0.84, P = .38$). However, from the descriptive plot in **Figure 6.8**, we can see the different trends of the two groups: A decrease of sedentary hours in Group A and an increase in Group B. The result of Bayesian paired samples $t$-test suggested (with weak evidence) that the daily sedentary hours in baseline week were larger than those in intervention weeks in group A (BF$_{+0} = 1.922, M = 0.522, CI = [0.044, 1.251]$). By contrast, in group B, it was more likely that the intervention had no effect than a positive effect with moderate evidence (BF$_{+0} = 0.278, M = 0.175, CI = [0.007, 0.644]$).

![Figure 6.8](image.png)

**Figure 6.8:** The participants’ daily sedentary hours based on the app-logged data during the baseline week and the intervention weeks. The bars refer to the confidence intervals with 95% confidence level.
Therefore, regarding RQ2, the intervention involving the visualizations and action planning in SedVis had a positive effect on reducing participants’ sedentary hours with weak evidence. Meanwhile, action planning alone had no effect on reducing participants’ sedentary hours with moderate evidence.

6.4.5 RQ3: Participants’ Interaction with SedVis

The frequency of checking the visualizations per day reflects the participants’ strength of self-monitoring, which might also act as a cue for self-reminding of sedentary behavior change. **Figure 6.9** shows the daily frequency of participants checking the visualizations in SedVis. We found that the participants were more likely to check the visualizations from the notification bar (68.33%) than from the dashboard (31.67%).

![Check Daily Sedentary Behavior Per Day](image)

**Figure 6.9:** The daily frequency of participants checking the visualizations in SedVis through the notification and the home screen.

To test the assumption that the more the participants engage with the app the more the app will affect their behavior, we correlated the daily frequency of participants checking the visualization in SedVis with their change of sedentary hours, calculated as daily sedentary hours during the intervention weeks minus the counterparts during the baseline week (see **Figure 6.10**). A Pearson correlation did not show a statistically significant correlation between participants’ checking the visualizations in SedVis with the change of daily sedentary hours \( r = -0.498, P = .21 \), although the effect was large (Cohen, 1992). Then we conducted a Bayesian Pearson correlation test with the alternative hypothesis of negative correlation. The Bayesian factor \( \text{BF}_{-0} = 1.489, r = -0.496 \) weakly suggested that the two factors were more likely to be negatively related than unrelated. To some extent, the result suggested that the participants’ engagement was positively related to the effect of reducing sedentary hours.
6.4.6 RQ4: User Experience

User experience was investigated both quantitatively and qualitatively. By comparing to the benchmark provided by the UEQ toolkit (Laugwitz, Held, & Schrepp, 2008), we mapped the participants’ scores of user experience to quality levels, as shown in Table 6.3.

Table 6.3: The user experience scores based on UEQ.

<table>
<thead>
<tr>
<th>UEQ Scales</th>
<th>Mean (Group A)</th>
<th>Comparison to Benchmark</th>
<th>Mean (Group B)</th>
<th>Comparison to Benchmark</th>
<th>t₁₄</th>
<th>P</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attractiveness</td>
<td>1.65</td>
<td>Good</td>
<td>1.44</td>
<td>Above average</td>
<td>0.60</td>
<td>.56</td>
<td>0.3</td>
</tr>
<tr>
<td>Perspicuity</td>
<td>2.25</td>
<td>Excellent</td>
<td>2.10</td>
<td>Excellent</td>
<td>0.47</td>
<td>.64</td>
<td>0.24</td>
</tr>
<tr>
<td>Efficiency</td>
<td>1.91</td>
<td>Excellent</td>
<td>1.84</td>
<td>Excellent</td>
<td>0.19</td>
<td>.85</td>
<td>0.10</td>
</tr>
<tr>
<td>Dependability</td>
<td>1.34</td>
<td>Above Average</td>
<td>1.22</td>
<td>Above Average</td>
<td>0.32</td>
<td>.75</td>
<td>0.16</td>
</tr>
<tr>
<td>Stimulation</td>
<td>1.66</td>
<td>Good</td>
<td>1.16</td>
<td>Above Average</td>
<td>1.72</td>
<td>.11</td>
<td>0.86</td>
</tr>
<tr>
<td>Novelty</td>
<td>1.22</td>
<td>Good</td>
<td>0.44</td>
<td>Below Average</td>
<td>2.75</td>
<td>.02</td>
<td>1.37</td>
</tr>
</tbody>
</table>

According to the results of Bayesian t-test with the alternative hypothesis that Group A scored higher than Group B, our visualizations yielded more perceived stimulation ($BF_{10} = 1.989$, $M = 0.621$, CI = [0.045 1.622]) and novelty ($BF_{10} = 7.439$, $M = 1.025$, CI = [0.160, 2.163]). For other aspects, the scores tended to be equivalent. We observed that the previewed dependability is only “above average,” which means the participants did not think the data shown in the app were very accurate. According to the exit interview, several participants believed their steps were underestimated based on two reasons: (1) the sensors in some
smartphones were not very sensitive; (2) they did not take the smartphone during some indoor activities, e.g., going to the restroom or walking in the laboratory.

Even though participants were asked to make at least one action plan every day during the two-week intervention phase, the average number of daily action plans was only 0.59, which hints that participants might not have used the app regularly. Regarding user acceptance, no participant complained about interruptions of daily activities through using the app, while some participants commented that making action plans every day was boring. Eight participants wanted to keep the app and continue using it for reducing sedentary behavior, while four participants implied they needed a reminder if they continued to use the app. Four participants said they did not want to continue the app for the following reasons: (1) “it underestimates my steps”; (2) “I do not want to always keep the GPS on”; (3) “the app provided too little new information”; (4) “I need a reminder for enacting my plans.”

6.5 Discussion

6.5.1 Principle Findings

In this chapter, we tested SedVis, an app-based sedentary behavior intervention which aims to reduce sedentary behavior through a combination of mobility pattern visualization and daily action planning. Compared to a control group that only received action planning, the intervention group slightly reduced their sedentary time. Contrary to our expectation, however, the visualizations did not impact on the participants’ action planning. As we observed an association between the change of sedentary time and the participants’ engagement with SedVis, we assume that the effect of SedVis on sedentary behavior might have been induced by self-monitoring, which might have been strengthened by the stimulation and novelty of our visualizations.

In line with Maher and Conroy (Maher & Conroy, 2015), the present study found no effect of daily action planning on reducing sedentary behavior in the short term among college students. Furthermore, data showed that sedentary behavior change did not correlate with the number of total action plans, the number of unique action plans, the perceived viability of action plans, and the specificity of action plans. As explained by Maher and Conroy (Maher & Conroy, 2015), one reason of the ineffectiveness could be that the cue-to-action response expected by action planning relies much on conscious self-regulatory process, which is difficult for highly habitual behavior, such as sedentary behavior. We would like to add another explanation using prospective memory (McDaniel & Einstein, 2007) inspired by the work of Grundgeiger and colleagues (Grundgeiger et al., 2017): Prospective memory tasks, which require us to remember to do something at a future time, are very difficult especially when we are focusing on other tasks. Because sedentary behavior is always coupled with other tasks holding our
attention, the action plans of reducing sedentary behavior might be easily forgotten.

The participants’ evaluations of SedVis with visualizations were good or excellent regarding the attractiveness, perspicuity, efficiency, stimulation, and novelty. Only the perceived dependability was at the level of above average. This may reflect some participants’ concern that SedVis underestimated their steps. At the exit interview, several participants mentioned that they believed the app missed part of their daily steps because they did not take the smartphone for some certain activities (e.g., working in the laboratory). This limitation of the present study could be avoided in future studies by using wearable sensors (e.g., wristbands, posture monitors, see (Atkin, Gorely, Clemes, & Yates, 2012)).

6.5.2 Implications for Future Work

6.5.2.1 Rethinking the action plans

While most participants made action plans adhered to the format of specifying “when”, “where”, and “how” to reduce their sedentary time, one participant additionally enclosed other contextual cues in their plans. For example, “15:00, lab, take a walk in between experiments” and “13:00, uni, walk between lectures.” Because of the additional cues—experiment and lectures here—the plans might be easier to remember. These plans are in line with the “if-then” format of implementation intentions, which emphasize the contextual cues linking to the goal-directed behavioral response (Gollwitzer, 1999; Hagger & Luszczynska, 2014). As sedentary behavior is prevalent, the cues of “When” and “Where” provide limited strength of conditional links to the response behavior. Due to the requirement of less self-regulatory resources, the more contextual plans in “if-then” format might be more effective than the plans in “when, where, and how” format (Hagger & Luszczynska, 2014; Maher & Conroy, 2015). However, no prior studies have compared the difference in sedentary behavior change. Relating to SedVis, future work might explore how the app could support personalized implementation intentions and its effectiveness on sedentary behavior change. For example, generating recommendations of plans based on users’ mobility patterns and context, which they might not even notice. Several heuristic rules could be used, e.g., going to the restroom downstairs instead of the nearest one, or more frequently going to the kitchen to drink water. Armitage (Armitage, 2009) found that experimenter-provided and self-generated implementation intentions could be equally effective in reducing alcohol consumption. It worth investigating this effect on sedentary behavior change following our study design. Some participants commented that making plans every day was boring, so generating plan-recommendations might also increase user acceptance in the long term.
6.5.2.2 Rethinking Self-Monitoring and Reminders

Since our study suggests that higher interaction frequency could lead to more reduction of sedentary behavior, future work might study the more convenient and intuitive user interfaces (e.g., glanceable (Gouveia, Pereira, Karapanos, Munson, & Hassenzahl, 2016)) to simplify the means of self-monitoring and interaction with the app even further. In the current version of SedVis, the easiest way to see the daily visualization was to swipe down the notification bar and click on the notification. In a future version, the app could illustrate the temporal sedentary information using an always-on progress bar (see details in Chapter 7) (Y. Wang & Reiterer, 2019) embedded in the notification or the app widget on the smartphone’s home screen. Future work should also considerate users’ need for reminders. Participants expressed different attitudes towards reminders: Some of them believed reminders for the action plans they made would be helpful because they sometimes forgot the plans; others thought that reminders would be unnecessary because of the potential interruption. Although fixed-time reminders (e.g., prompts on PC screens) were frequently used in prior interventions for reducing sedentary behavior at work (as shown in Chapter 2) (Y. Wang, Wu, et al., 2018), no studies explored the effectiveness and user experience of the reminders for personalized action plans.

6.5.3 Limitations

The present study contained several limitations. First, the sedentary hours based on the app-logged data might underestimate the participants’ movements. One reason for this might be that participants might not take the smartphone with them during certain activities such as going to the bathroom. Another reason could be that some activities could not be recognized and counted as steps. For example, one participant made an action plan to do push-ups at home, which cannot be recognized and recorded using a smartphone’s sensors. Second, the sample size is small. The small sample size and the relatively large between-subjects variances of the measurements might be the reason for several weak conclusions. For example, we observed that the means of the number of action plans, the number of unique action plans, and specificity (“When” and “Where”) were higher in Group A than in Group B. But only statistically weak evidence could be found. Lastly, the study period is relatively short, which limits the validation of our results in short-term scenarios. Therefore, future studies should replicate the present results in larger samples and with longer study duration.

6.6 Conclusion

Mobile apps on smartphones hold great potential for sedentary behavior change. Using novel visualizations of mobility patterns might be a promising avenue for reducing sedentary time by facilitating self-monitoring of the behavior and providing engaging feedback.
7 SUPPORTING SEDENTARY BEHAVIOR CHANGE USING CONTEXT-AWARE REMINDERS ON PC SCREEN

好雨知时节，当春乃发生。随风潜入夜，润物细无声。

Propitious rain comes opportuneiy in spring, slipping into night with breeze, gratifying the thirst of life silently.¹

- 杜甫 (712 ~ 770)
DU Fu
(A prominent Chinese poet of the Tang dynasty)
¹The translation is adapted from the version by Xueqing Zhang (章学清).


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Harald Reiterer supervised the work and reviewed this paper.
7.1 Abstract

An appropriate reminder could help screen-based workers to reduce their prolonged sedentary behavior. The fixed-duration point-of-choice prompt has been frequently used in related work. However, this prompting system has several drawbacks. In this chapter, we propose the SedBar, a context-aware reminding system using an always-on progress bar to show the duration of a working session, as an alternative to the prompt. The new reminding system uses both users’ keyboard/mouse events on the computer and the state-of-the-art computer vision algorithm to detect users’ presence from the webcam, which makes the system more accurate and intelligent. Our evaluation study compared the SedBar and a prompt based on the same context-aware system using both subjective and objective measures. After using each method for a week respectively, more participants preferred the SedBar. The participants’ perceived interruption and usefulness also suggested the SedBar was more popular during the study. However, the objective measurements of the participants’ working sessions indicated only the prompt was significantly effective in reducing their sedentary behavior.

7.2 Introduction

For screen-based workers (e.g., office workers and college students), prolonged sedentary behavior is ubiquitous, prevalent, and routine. Both academia and industry have drawn attention to technologies to help users to reduce prolonged sedentary behavior. In Chapter 2, our systematic review on persuasive technology in reducing prolonged sedentary behavior at work revealed that the fixed-duration point-of-choice prompt with motivational messages on PC was the most commonly used method among the reviewed empirical studies (Y. Wang, Wu, et al., 2018).

In a recent CHI paper, Luo and colleagues (Luo et al., 2018) reported their research on how screen-based workers interacted with a break prompting system on PC. The authors conducted a field study with 25 participants for three weeks. The participants responded 74% of the prompts, while 46% of the responses were “not to take a break at the moment.” They asked the participants to log the reasons if they did not take a break when they received a prompt. The user-logged information indicated eight reasons as shown in Table 7.1. These reasons implied two drawbacks of the prompt system: it is not aware of users’ working/breaking state (fixed-duration; see Reason 3); it does not allow users to prepare for a break (point-of-choice; see Reason 4 and 5). These drawbacks might increase users’ perceived interruption and decrease the user experience.
Table 7.1: Reasons for not taking breaks when receiving the prompts in (Luo et al., 2018).

<table>
<thead>
<tr>
<th>Reason</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 busy working</td>
<td>40.6</td>
</tr>
<tr>
<td>2 in a meeting or class</td>
<td>18.9</td>
</tr>
<tr>
<td>3 coming back from a break</td>
<td>15.2</td>
</tr>
<tr>
<td>4 close to finishing something</td>
<td>8.8</td>
</tr>
<tr>
<td>5 heading to other places (e.g., home) soon</td>
<td>7.9</td>
</tr>
<tr>
<td>6 engaged in a conversation</td>
<td>3.7</td>
</tr>
<tr>
<td>7 engaged in a screen-based activity (e.g., video, game)</td>
<td>2.5</td>
</tr>
<tr>
<td>8 having lunch or dinner</td>
<td>2.4</td>
</tr>
</tbody>
</table>

To solve the mentioned problems of fixed-duration point-of-choice prompts, we developed a new reminding system running on users’ working computers. We used the keyboard/mouse events and the webcam with the state-of-the-art deep learning algorithms to recognize users’ presence (working/breaking state). For the visualization of reminders, we implemented an always-on progress bar to indicate the current working (sedentary) duration and a conventional prompt. In this chapter, we aim to answer two research questions:

(RQ1) Which reminder is more effective on reducing sedentary behavior of screen-based workers?

(RQ2) Which reminder do users prefer?

7.3 Related Work

We searched the ACM digital library to collect related work on technologies of sedentary behavior detection and intervention (see Table 7.2). Among the 11 listed papers from 2013 to 2018, smartphones, sit pads, computers, and extra motion trackers were used to detect users’ sedentary behavior. We think the working computer (desktops/laptops) is the best platform for sedentary behavior change intervention because: (1) screen-based workers focus on their working computers most of the time; (2) extra devices - including smartphones - might add unnecessary distraction or setup burdens to users. Therefore, we use the keyboard/mouse and the webcam as the context detection tools, while we use a screen widget as the intervention cue.

We base our design on some prior work. In 2014, Mateevitsi et al. (Mateevitsi et al., 2014) proposed the HealthBar, an ambient persuasive device that helped users to break up their prolonged sedentary habits. The HealthBar used a passive infrared motion sensor to detect users’ presence/absence from their working desks. It used a three-foot plastic diffuser light-tube to provide feedback to users. The color of the HealthBar changed along with users’ working duration. The authors conducted a five-day pilot study with eight office workers to evaluate
the HealthBar. Qualitative results showed that the HealthBar could be a non-distracting and effective solution for reducing the sedentary behavior of office workers. However, the hardware setup of the HealthBar might hinder its deployment in large-scale use. Therefore, we use a screen widget instead of a physical light-tube as the reminder (see Section 3.2 for details).

**Table 7.2: Related work.**

<table>
<thead>
<tr>
<th>Paper</th>
<th>Platform</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Hirano et al., 2013)</td>
<td>Smartphone</td>
<td>A pilot field study (N=8, four weeks)</td>
</tr>
<tr>
<td>(S. J. Wang &amp; Yu, 2013)</td>
<td>Sit pad</td>
<td>None</td>
</tr>
<tr>
<td>(van Dantzig et al., 2013)</td>
<td>Smartphone; accelerometer-based motion tracker; computer</td>
<td>A feasibility study (N=8, one day); An evaluation study (86 participants, seven weeks)</td>
</tr>
<tr>
<td>(Mateevitsi et al., 2014)</td>
<td>Infrared motion sensor</td>
<td>A feasibility study (N=8, five days)</td>
</tr>
<tr>
<td>(Ferreira et al., 2014)</td>
<td>Computer with webcam</td>
<td>None</td>
</tr>
<tr>
<td>(Min et al., 2015)</td>
<td>Sit pad; smartphone</td>
<td>None</td>
</tr>
<tr>
<td>(van Schagen et al., 2015)</td>
<td>Smartphone</td>
<td>None</td>
</tr>
<tr>
<td>(Pinder et al., 2015)</td>
<td>Smartphone</td>
<td>None</td>
</tr>
<tr>
<td>(Grundgeiger et al., 2017)</td>
<td>Smartphone</td>
<td>A pilot study (N=5, five days)</td>
</tr>
<tr>
<td>(Wölfel, 2017)</td>
<td>Kinect</td>
<td>A feasibility lab study (N=16, 3 hours)</td>
</tr>
<tr>
<td>(Luo et al., 2018)</td>
<td>Computer</td>
<td>An exploratory field study (N=25, three weeks)</td>
</tr>
</tbody>
</table>

Also in 2014, Ferreira et al. (Ferreira et al., 2014) presented the BreakOut, a desktop system aiming to infer users’ posture, stress level, and engagement with computer-related tasks for recommending breaks at the appropriate time. However, there is no following intervention study to evaluate its effectiveness. The BreakOut used the keyboard/mouse events and the webcam to detect users’ working engagement and posture simultaneously. Using the same hardware setting, we detect users’ presence by combining advanced computer vision algorithms and the keyboard/mouse events in a more efficient way.

As should be noticed, only one (van Dantzig et al., 2013) of the reviewed work listed in Table 7.2 used control studies to evaluate their intervention approach. The others focused on the system description and validation. To evaluate our proposed system, we conducted a three-week field study (See Section 7.5 for details).
Chapter 7: Supporting Sedentary Behavior Change Using Context-Aware Reminders on PC Screen

7.4 Intervention System

7.4.1 Context Detection

The awareness of users’ presence could help the system to decrease unnecessary reminders. E.g., if a user takes a break before the prompt appears, the system should reset the timer to avoid the surplus reminding (Reason 4 in Table 7.1). Only using the keyboard/mouse events to determine working state is efficient, but could be inaccurate when a user reads documents or watches videos that require no keyboard/mouse interaction. Therefore, we also use the webcam to detect users’ presence. We adopt the state-of-art person-detection and face-detection computer vision algorithms based on the deep neural network (DNN) module in OpenCV library. Detecting the user’s presence using the computer vision algorithm requires much more computation load than the using the keyboard/mouse events. Therefore, we combine the two methods: we only run the computer vision algorithm when there are no keyboard/mouse events for half a minute; if the user is absent for the following half minute, the state changes to breaking (see Figure 7.1). The computer vision algorithm does not run in the breaking state. No video or picture will be recorded for privacy consideration.

7.4.2 Intervention Cues

A point-of-choice prompt (see Figure 7.2) on PC is usually an alert window with a short message showing up for a short period, which was frequently used in related work. Upon receiving a prompt, a user is expected to decide on whether taking a break or not. One drawback of the prompt is that it does not allow users to prepare for breaks because it is invisible until the pre-defined time is up. Also, the sudden appearance of the prompt might cause users’ pressure and distraction. If a user is busy (Reason 1 in Table 7.1) when receiving a prompt, it could also be a user burden to make a decision.

Figure 7.1: The state transitions between working and breaking. W-B condition: a user is absent for more than a minute in the working state. B-W condition: a user starts to use the mouse or the keyboard in the breaking state.
Figure 7.2: The prompt interface. In our study, it shows up when the current duration of working state exceeds the pre-defined duration. It disappears when the user clicks on the button or after 30 seconds without interaction.

Figure 7.3: The SedBar. In the working state, the progress bar color is orange by default, while the progress indicates the working duration. In the breaking state, the progress bar color is blue by default, while the progress loops like a battery charging animation.

As an alternative, an always-on progress bar provides some superior features. We call it the SedBar, as shown in Figure 7.3. Instead of the sudden appearance, the SedBar is always on and glanceable, thus avoiding the time pressure of decision-making when using a prompt. Being aware of the working duration, a user has enough time to adjust the work and prepare for a break. It could be more interruptive than the prompt because it is always present. It could also be less interruptive because the progress bar grows so slowly that a user might not notice when it is full (time is up). Therefore, we need to study both their effectiveness and interruption.

7.5 Study Design

To compare the always-on progress bar with the point-of-choice prompt, we conducted a three-week field study with eight participants in the University of Konstanz during November and December in 2018. We adopted a single-case design in this study (Dallery, Cassidy, & Raiff, 2013), as shown in Figure 7.4. We assigned the participants to two groups alternatively according to the order they contacted us for the study. We deployed the two study conditions – Prompt and SedBar – in opposite sequence to balance the potential novelty effect of using the software. At appointment 1 (the first blue dot in Figure 7.4), the participants read the information sheet, signed the informed consent, and installed the software on their computers. At appointment 2, we showed them an educational video and a
flyer to explain the potential health problems caused by the sedentary lifestyle. At appointment 4, we interviewed the participants and asked them about their preferences between the two intervention cues (reminders).

During the baseline week, the software ran in the background and no intervention cue was shown to the participants. During the following two weeks of interventions, we asked them to fill two four-point scales about their perceived usefulness (from “not useful at all (1)” to “very useful (4)” and interruption (from “not interruptive at all (1)” to “very interruptive (4)” of the reminder once a day.

![Figure 7.4: The study procedure. The dots represent the appointments during the study. The SedBar and the Prompt use the same context detection method and follow the state transition conditions as shown in Figure 7.1. The dots refer to the appointments.](image)

During the interventions conditions, each participant set a reminder duration of 30-60 minutes and could adjust it during the study. For the SedBar, the participants could also change its position, size, and color. We allowed them to stop the software whenever they wanted. Our software logged the participants’ usage periods, state transitions, and settings. The software ran offline - all the data were stored locally. We collected the data by email from the participants at each appointment.

To conclude, the measures in the study included the participants’ work sessions (i.e., times and durations), daily perceived usefulness and interruption, and their preferences of the reminders.

Eight college students participated in this study (five Ph.D. student; two master students; one bachelor student); two of them were male. All the participants had the intention to reduce their sedentary behavior and were neither physically disabled nor ill during the study.

### 7.6 Statistical Analysis

Following the statistical method in the SedVis study (see Chapter 6 for details), we used both conventional null-hypothesis significance testing (i.e., t-tests and ANOVAs) as well as Bayes factors. The analysis tool and the reporting method were also the same as in the SedVis study. The single-case study design allowed us to analyze each participant’s measurements in the two interventions conditions comparing to the baseline condition. Besides, we aggregated all the participants’ data in each condition to analyze the general effects of each intervention condition.
7.7 Results

7.7.1 Prolonged Sedentary Sessions

We regarded the working state logged by our software as the sedentary state of the participants because no participant used a standing desk during the study. Figure 7.5 shows the results of the one-way ANOVA and the post hoc test of the prolonged sedentary sessions (t > 30 minutes) for each participant. For three participants (A3, B2, and B4), the decrease of the sedentary duration in the Prompt condition was statistically significant. For most of the participants, the prompt seemed to be more effective than the SedBar. We found no statistical significance between the SedBar condition and the control condition.

![Figure 7.5: The results of post hoc tests for the prolonged work sessions (t > 30 minutes) in three conditions for each participant. The horizontal axis represents duration in minutes. The red bar indicates the statistically significant difference compared to the selected condition (the blue bar).](image)

The result of a one-way ANOVA showed that there was a significant effect of our intervention on participants’ prolonged work sessions for the three conditions ($F_{2, 421} = 20.09, P < .001$). Post hoc comparisons using the Tukey’s test indicated that the mean durations of prolonged work sessions for the Prompt condition (mean = 40.20, SD = 8.65) was significantly less than the baseline condition (mean = 59.41, SD = 28.55). By contrast, no significant difference was found in the SedBar condition (mean = 55.94, SD = 26.47) compared to the baseline condition, as shown in Table 7.3. A Bayesian ANOVA and the corresponding post hoc tests confirmed the effect in the Prompt condition with very strong confidence ($BF_{10} = 2.12 \times 10^7$) and indicated no effect in the SedBar condition with moderate...
confidence ($BF_{10} = 0.13$) compared to the baseline condition. Figure 7.6 visually illustrates the differences among the three conditions.

Table 7.3: The results of post hoc tests for the durations of prolonged work sessions.

<table>
<thead>
<tr>
<th></th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>$t$</th>
<th>Cohen $d$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Prompt</td>
<td>19.22</td>
<td>3.136</td>
<td>6.128</td>
<td>0.835</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Baseline SedBar</td>
<td>3.49</td>
<td>2.713</td>
<td>1.286</td>
<td>0.127</td>
<td>.40</td>
</tr>
<tr>
<td>Prompt SedBar</td>
<td>-15.73</td>
<td>3.096</td>
<td>-5.080</td>
<td>-0.726</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Figure 7.6: The results of post hoc tests for the prolonged work sessions ($t > 30$ minutes) in three conditions aggregated of all participants. The horizontal axis represents duration in minutes. The red bar indicates the statistically significant difference compared to the selected condition (the blue bar).

Regarding the numbers of the participants’ work sessions, the result of a repeated measures ANOVA did not show any significant difference among the conditions ($F_{2, 421} = 3.69$, $P = .052$). Figure 7.7 visually illustrates the differences among the three conditions. Because the $p$ value was close to the significance threshold, we conducted a post hoc test, which indicated a significant effect in the Prompt condition (mean = 12.38, SD = 6.19) compared to the baseline condition (mean = 19.63, SD = 5.48), as shown in Table 7.4. A Bayesian repeated measures ANOVA and the corresponding post hoc tests suggested the effect in the Prompt condition with moderate evidence ($BF_{10} = 5.31$) and indicated no effect in the SedBar condition with weak evidence ($BF_{10} = 0.36$) compared to the baseline condition.

Table 7.4: The results of post hoc tests for the number of prolonged work sessions.

<table>
<thead>
<tr>
<th></th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>$t$</th>
<th>Cohen $d$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Prompt</td>
<td>7.250</td>
<td>2.194</td>
<td>3.305</td>
<td>1.168</td>
<td>.04</td>
</tr>
<tr>
<td>Baseline SedBar</td>
<td>-1.375</td>
<td>3.580</td>
<td>-0.384</td>
<td>-0.136</td>
<td>1.00</td>
</tr>
<tr>
<td>Prompt SedBar</td>
<td>-8.625</td>
<td>4.157</td>
<td>-2.075</td>
<td>-0.733</td>
<td>.23</td>
</tr>
</tbody>
</table>
7.7 Results

Figure 7.7: The results of post hoc tests for the numbers of the prolonged work session ($t > 30$ minutes) in three conditions aggregated of all participants. The horizontal axis represents duration in minutes.

7.7.2 Perceived Interruption and Usefulness

The scores of the perceived interruption and usefulness are shown in Figure 7.8 and Figure 7.9, respectively. We compared the scores of the SedBar condition and the prompt condition for each participant: the smaller scores the better for the perceived interruption; the larger scores the better for the perceived usefulness. In Figure 7.8, we can see that (1) most of the participants. We observe mixed patterns from the data. More participants thought the SedBar is less interruptive and more useful. Only two participants perceived more interruption of the SedBar, while three participants thought the prompt was more useful to them.

Figure 7.8: The bar chart (with the error bar of standard deviation) of each participant’ perceived interruption for SedBar and the prompt. The smaller the better. The solid blue bars highlight the cases where the SedBar was perceived less interruptive than the prompt.
Figure 7.9: The bar chart (with the error bar of standard deviation) of each participant’s perceived usefulness for SedBar and the prompt. The larger the better. The solid blue bars highlight the cases where the SedBar was perceived more useful than the prompt.

The result of an independent samples t-test (one-sided) showed that there was a significant difference for perceived usefulness between the two conditions ($t_{70} = -1.73$, $P = .04$, Cohen $d = -0.41$), while there was no significant difference for perceived interruption between the two intervention conditions ($t_{70} = -0.22$, $P = .41$, Cohen $d = -0.05$), as shown in Figure 7.10. We then conducted a Bayesian independent samples t-test (one-sided), which suggested the perceived usefulness in the SedBar condition was higher than in the Prompt condition with weak evidence ($\text{BF}_{0} = 0.61$, $M = -0.37$, CI = [-0.81, -0.04]) and the perceived interruption was the same in two conditions with moderate evidence ($\text{BF}_{0} = 3.46$, $M = -0.16$, CI = [-0.53, -0.01]).

Figure 7.10: The bar chart (with Confidence Interval) of the perceived interruption and usefulness aggregated from all the participants in the two intervention conditions.

7.7.3 Participants’ Preferences and Reasons

In the final interview, seven participants confirmed they would like to continue to use the software. In group A, two participants preferred the prompt, while the
other two preferred the SedBar. In group B, one participant preferred the prompt, while the others preferred the SedBar. Overall, the SedBar was more popular than the prompt for the following reasons: (1) it allowed preparation for breaks (A2); (2) it provided more information (A3); (3) it was always visible (B1, B3 and B4); (4) it provided stronger intervention (B3); (5) it was more interesting (B4). The reasons why some participants chose the prompt included: (1) it was less annoying (A1 and B2); (2) the SedBar covered some content on the screen (A4); (3) the state-change of the SedBar was inaccurate sometimes (A4). Table 7.5 shows all the participants’ preferences and their reported reasons.

Table 7.5: The participants’ ratings on the reminders, their preferences, and reported reasons. PI refers to the perceived interruption; PU refers to the perceived usefulness. The blue parts indicate the cases when the SedBar was preferred and the assumed attributing ratings in these cases.

<table>
<thead>
<tr>
<th></th>
<th>SedBar</th>
<th>Prompt</th>
<th>Preference</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>2.8</td>
<td>2.8</td>
<td>2.3</td>
<td>Prompt</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
<td>4</td>
<td>1.4</td>
<td>SedBar</td>
</tr>
<tr>
<td>A3</td>
<td>2</td>
<td>3.2</td>
<td>2.8</td>
<td>SedBar</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
<td>3</td>
<td>1.5</td>
<td>Prompt</td>
</tr>
<tr>
<td>B1</td>
<td>1</td>
<td>3.2</td>
<td>1.3</td>
<td>SedBar</td>
</tr>
<tr>
<td>B2</td>
<td>1</td>
<td>3</td>
<td>1.3</td>
<td>Prompt</td>
</tr>
<tr>
<td>B3</td>
<td>1.7</td>
<td>3.2</td>
<td>1</td>
<td>SedBar</td>
</tr>
<tr>
<td>B4</td>
<td>1</td>
<td>2</td>
<td>1.5</td>
<td>SedBar</td>
</tr>
</tbody>
</table>

7.8 Discussion

7.8.1 Principle Findings

There was a disparity between the objectively measured sedentary behavior change and the user-perceived usefulness: the prompt helped participants reduce the number and duration of their prolonged work sessions with statistical significance while the SedBar did not show significant effects (RQ1); the perceived usefulness for the SedBar was higher than that for the prompt with statistical significance and more participants preferred the SedBar at the end of the study (RQ2). Although the SedBar did not show statistically significant
effects on sedentary behavior change, the participants still perceived that it was useful. Future work should investigate the underlying mechanism of users’ preference. We found that the perceived interruption in both conditions was relatively low. Therefore, future work could consider more obtrusive features for larger effects on the behavior change. Besides, it is worthy of investigating how to combine these two methods for better reminding users of breaking prolonged sedentary behavior in future work.

7.8.2 Personalization

The context detection in our intervention system avoided the surplus reminding, thus increasing personalization to some extent, which might be related to the low perceived interruption. The participants’ preferences and the reported reasons in the final interview suggest that our design concept of the SedBar is meaningful in reality: it does not push users when the time is up; it allows users to prepare for breaks. Besides, we found more opportunities to improve personalization, including adjusting the size of the SedBar to avoid covering the content on users’ screen and considering the multi-screen case for some users.

7.8.3 Limitations

Due to the small user size and short study duration, further studies are required to compare the effectiveness and user experience of the two reminding methods.

7.9 Conclusion

Context-aware PC reminders hold great potential for sedentary behavior change. Using the work state detection based on keyboard/mouse events and human/face recognition by webcam, the point-of-choice prompt might be a promising tool for reducing sedentary time for screen-based workers. Users perceived the always-on progress bar as useful, but the effectiveness should be further investigated.
7.9 Conclusion
Chapter 8: Exploring the Impact of Augmented-Reality Head-Mounted Displays on Users’ Movement Behavior

8 EXPLORING THE IMPACT OF AUGMENTED-REALITY HEAD-MOUNTED DISPLAYS ON USERS’ MOVEMENT BEHAVIOR

君子以思患而豫防之。

Nip in the bud.

- 《易经》
  *Book of Changes*
  (The oldest of the Chinese classics)


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Harald Reiterer supervised the work and reviewed this paper.
8.1 Abstract

The augmented-reality head-mounted display (e.g., Microsoft HoloLens) is one of the most innovative technologies in multimedia and human-computer interaction in recent years. Despite the emerging research of its applications on engineering, education, medicines, to name a few, its impact on users’ movement behavior is still underexplored. The movement behavior, especially for office workers with sedentary lifestyles, is related to many chronic conditions. Unlike the traditional screens, the augmented-reality head-mounted display (AR-HMD) could enable mobile virtual screens, which might impact on users’ movement behavior. In this chapter, we present our SedHolo study to explore the impact of AR-HMDs on users’ movement behavior. We compared the differences of macro-movements (e.g., sit-stand transitions) and micro-movements (e.g., moving the head) between two experimental modes (i.e., spatial-mapping and tag-along) with a dedicated trivial quiz task using HoloLens. The study reveals interesting findings: strong evidence supports that participants had more head-movements in the tag-along mode where higher simplicity and freedom of moving the virtual screen were given; body position/direction changes show the same effect with moderate evidence, while sit-stand transitions show no difference between the two modes with weak evidence. Our results imply several design considerations and research opportunities for future work on the ergonomics of AR-HMDs in the perspective of health.

8.2 Introduction

Augmented-reality head-mounted displays (AR-HMD, e.g., HoloLens) have been popular in both academia and industry. This multimedia platform enables novel interfaces between users and information, which brings immersive and in-situ user experience. However, the ergonomics, especially the health impacts of using these devices, is still underexplored.

Since the beginning of the PC era, human labor work has been increasingly replaced by machines and computers. Inactive office work becomes many people’s lifestyle. However, human’s body cannot evolve so fast as the modern industry that many chronic diseases become pervasive. In the short term, static postures that cause much pressure to our spines could contribute to musculoskeletal disorders (e.g., back and neck pain) (Grandjean & Hünning, 1977). In the long term, inactive lifestyles are related to obese (Fountaine, Piacentini, & Liguori, 2014), type 2 diabetes (van der Ploeg et al., 2012), cardiovascular diseases (Warburton & Bredin, 2016), and even certain cancers (e.g., colon cancer (Howard et al., 2008)). Recent studies (Knaeps et al., 2016) showed that even meeting the well-established physical activity (PA) recommendations (i.e., 150-300 minutes of moderate intensity PA or 75-150 minutes of vigorous intensity PA per week) cannot compensate the bad effect of sedentary behavior (prolonged-sitting) at work.
Therefore, improving individuals’ inactive work style is urgent. Both academia and industry have studied many methods - from changing working policies and physical environments to providing dedicated education and reminders (Gardner et al., 2016). According to Fogg’s Behavioral Model (FBM) (Fogg, 2009a), human behaviors are determined by three factors: ability, motivation, and triggers. Lacking one of these factors will lead to the failure of the target behavior. In Chapter 2 (Y. Wang, Wu, et al., 2018), we illustrated that education plus point-of-prompt PC reminders could be effective in reducing sedentary behavior at work. This mixed intervention strategy enhances two factors in FBM: motivation and triggers. Regarding the third factor, ability, changing workplace settings or adding facilities (e.g., using sit-stand work station or treadmill) could be helpful. One way to increase users’ ability is to make the target behavior easier to do, which is the focus of this chapter.

Despite the approaches to solving the inactivity problem in office work, we have to ask what causes the inactive lifestyle. The reasons could be multiple: the limited room space, being quiet to avoid interrupting colleagues, focusing on tasks, and so on. However, we believe the fixed computer screen is a non-negligible factor. To visually obtain the information from the computer, we have to stay close to the screen. Even using a sit-stand work station with multiple screens, users’ postures are still restricted by the screen’s position and size. Therefore, allowing users to move the screen freely could be a way to encourage their movements during work, which corresponds to the factor of ability in FBM. The AR-HMD (e.g., HoloLens) could enable free movements with the virtual screen in the device.

Therefore, we are eager to explore the potential impact of AR-HMDs on users’ movement behavior at office work, which could be applied to improve the screen-based office workers’ health. A representative of AR-HMDs is Microsoft HoloLens. It enables the experience of holograms, which is a new visual media with high potentials in human-computer interaction. Four features of HoloLens could meet the requirements of our study design: self-contained (wireless), spatial-awareness, movement-awareness, and augmented reality. In this chapter, we will present our exploratory study to answer the following research questions:

1. Will higher simplicity and freedom of moving the virtual screen in HoloLens lead to more movements of the users?
2. What are the main factors affecting users’ movements when using the virtual screen in HoloLens?

8.3 Related Work

AR-HMDs combine the advantages of two technologies: augmented reality (AR) and the head-mounted display (HMD). The emerging research of AR-HMDs has enabled applications on education (M. Wang, Callaghan, Bernhardt, White, &
Peña-Rios, 2018), engineering (Syberfeldt, Danielsson, & Gustavsson, 2017), medicine (L. Chen, Day, Tang, & John, 2017), and so on (Cipresso, Giglioli, Raya, & Riva, 2018; Patterson, Winterbottom, & Pierce, 2006). Although increasing work is focused on novel applications of AR-HMDs, the study of their impacts on human health is limited.

Yuan and colleagues (Yuan J, Mansouri B, Jh, Sf, & Khaderi, 2018) systematically reviewed the HMD-caused visual discomfort, indicating that the exposure to HMDs resulted in higher visual discomfort (i.e., simulator sickness and visual strain) compared with exposure to traditional displays such as TV and desktop computer displays. However, this review only covered studies using virtual reality HMDs. There is no systematic review of the visual discomfort impact by AR-HMDs due to the lack of related empirical studies. In a recent paper, Vovk et al. (Vovk, Wild, Guest, & Kuula, 2018) conducted an experiment using HoloLens with 142 subjects in three different industries (i.e., aviation, medical, and space), finding that HoloLens causes only negligible symptoms of simulator sickness across all participants.

Focusing on users’ postures, recently, Aromma et al. (Aromaa, Väätänen, Kaasinen, Uimonen, & Siltanen, 2018) reported a study evaluating a tablet-based AR system in maintenance work regarding human factors and ergonomics. Their study showed that the users adopted varied kinds of postures, of which some postures may increase the risk of musculoskeletal disorders in the long term. However, this study has two limitations: the selected task only took 20-30 minutes; the maintenance task in the study largely determined the users’ selected postures. In other words, the study was too short while the task did not allow much freedom of users’ movements and postures. In our study, we address the limitations by designing dedicated virtual-screen based task using HoloLens.

Although the focus of this chapter is exploring the potential impact of AR-HMDs on users’ movement behavior, the goal behind is to improve users’ health in office work. Therefore, we also discuss some related work focused on posture monitoring/guiding and physical activity promotion in office work, which could inspire the ergonomics research of postures and movements using AR-HMDs.

In a recent paper (Wu et al., 2018), Wu and colleagues proposed ActiveErgo using sensors and an automatically adjustable screen to improve users’ sitting postures. The study of ActiveErgo showed that the participants need support to follow the ergonomics guidelines. However, although the right posture is helpful to reduce musculoskeletal pain, prolonged sitting or standing is still detrimental for health. Increasing evidence (Callaghan, De Carvalho, Gallagher, Karakolis, & Nelson-Wong, 2015; K. G. Davis & Kotowski, 2015) suggests that a dynamic work style – e.g., reducing sitting while increasing sit-stand transitions – is superior to only sitting or standing.

To promote physical activity at work, Probst presented the Active Office (Probst, 2015), where new interaction technologies were designed to enable more diverse
movements in office work than the traditional point-and-click interaction. For example, a user controls a PC through the movements of her/his body (e.g., tilting, rotating, bouncing) while sitting on an adjustable office chair with sensors. The proposed technologies in Active Office were focused on the interaction between users and fixed screens, which could still limit users’ movements during their office work. However, the idea of integrating more movements to the office work routines also applies in the ergonomics design for AR-HMDs.

8.4 Study Design

8.4.1 Device

Several AR-HMDs are available in the consumer market recently, e.g., Microsoft HoloLens\(^\text{18}\), Magic Leap One\(^\text{19}\), and HTC Vive Pro\(^\text{20}\). As introduced earlier, HoloLens is a self-contained AR-HMD, which was released by Microsoft in 2016. In our study, we use the HoloLens due to its rich development documents and its large community of AR-HMD research. HoloLens uses the head gaze to control the virtual cursor and the air-tap hand gesture (or a remote clicker) to select the virtual icons. It also has several limitations, e.g., the small field of view and the weight. We considered these limitations when we design the task for our study.

8.4.2 Task

Since we want to explore the effect of the simplicity and freedom of moving the virtual screen in HoloLens on users’ movement behavior, we need to design the task that minimizes the impact of other factors. In the review of Yuan and colleagues (Yuan J et al., 2018), they discussed the effect of HMDs’ optical characteristics (system features), participants’ gender (individual characteristics), task duration and content (task characteristics) on users’ visual discomfort. We believe these factors will impact on users’ movement behavior as well. Therefore, we considered these factors in our study design.

We used a trivial quiz on a 2D virtual screen as the user task in our study (see Figure 8.1). The quiz contains 600 hundred questions, which took around one hour to finish. We intently chose interesting questions to decrease the boringness and increase participants’ concentration during the study. This task is to simulate a simple office work with moderate cognitive load. The reason for choosing the task based on a 2D virtual screen instead of 3D objects is to avoid the effect of the 3D object on users’ movements. To answer one question in the quiz, a user should

\(^{18}\) https://www.microsoft.com/en-us/hololens

\(^{19}\) https://www.magicleap.com/magic-leap-one

\(^{20}\) https://www.vive.com/ca/product/vive-pro/
gaze at the answer button then press the remote clicker. Using the click instead of the air-tap gesture is to avoid the arm fatigue after long-period use.

**Figure 8.1: An question example in the trivial quiz.**

HoloLens has a 30-degree horizontal and 17-degree vertical field of view, which is much smaller than the human eyes’ field of view. Therefore, we made the virtual screen a bit smaller than the view area. Thus, it required slight head movement to select the buttons on the screen, which is to minimize the effect of the field of view on users’ head movements.

**Figure 8.2: The physical environment in the study.**

We used two modes – spatial-mapping (Microsoft, 2018b) and tag-along (Microsoft, 2018a) - to provide different levels of simplicity and freedom of moving the virtual screen. In the spatial-mapping mode, users can put the virtual screen to any position on the walls within a given area (around 4 x 4 meters) in a room (see **Figure 8.2**). The visual range to the virtual screen is from 0.85 to 3.1 meters away due to the limit of the depth-perception capacity in HoloLens. Differently, in the tag-along mode, the virtual screen automatically follows the user when it is out of the view frustum as the user moves. Otherwise, the virtual screen stays straight to the user’s viewpoint at two-meter away. Besides, the spatial-mapping only allows users to attach the virtual screen on the wall in the given room, while the virtual screen has not the limit in the tag-along mode. To conclude, both the modes allow users to control the virtual screen, but the tag-along mode provides more simplicity (automatic following) and freedom (any posture of the virtual screen).
There were a table and an ergonomic chair in the study room. The users could sit, stand, and walk using any postures during the one-hour task. As we aimed to observe users’ voluntary behavior, we did not use any indicators or reminders during the study.

8.4.3 Participants and Procedure

We conducted the study in February 2019. We recruited ten healthy adults (female=5, age=27.2±2.9) from our university using emails and social networks. All the participants had no experience of using HoloLens before the study.

Before the users started the task, we assessed the participants’ perceived habit strength using the self-report index of habit strength index (Verplanken & Orbell, 2003) and gave them a short introduction of the adverse effect of inactive lifestyle on health. We then told the participants they should focus on the quiz and choose the position and posture as they like during the study. They could see the score on the corner of the virtual screen during the study (See Figure 8.1). Then the participants signed the informed consent form, which told them the usage of the collected data and that we would give 20 euros to each participant as compensation after the study.

The study consisted of two appointments on two successive days for each participant. We randomly chose five participants to use the spatial-mapping mode at the first appointment and then the tag-along mode at the second appointment. The other five participants used the two modes in the opposite sequence. The questions in the quiz were randomized each time. The cross-over study design was to balance the potential novelty effect of using HoloLens on the two modes. Most of the participants attended the two appointments at the same time on the two successive days, which is to avoid the potential effect of the day time on the participants’ behavior. For example, a participant might prefer to sit down in the evening due to tiredness. Two participants were exceptional because of their schedule limitation. Therefore, we counterbalanced their study time for the two modes. After the quiz, we assessed participants’ workload using the NASA-TLX questionnaire (Hart & Staveland, 1988). The participants were allowed to drop the study or take a break if they feel uncomfortable when wearing the device during the study.

At the end of each appointment, we conducted a semi-structural interview with each participant, which took around 30 minutes. The questions in the interview included:

1. How did you adjust the positions of the virtual screen?
2. How did you choose your position and postures?
3. Which mode do you prefer? And why?
4. Which features would you like to improve or add in the future AR applications on HoloLens?
We used the last two questions only at the end of the second appointment. The interview provided us more insights about the participants’ behavior, in addition to objective measurements.

### 8.4.4 Measurements

We recorded the participants’ behavior during the study using two cameras to cover the study room. Besides the videos, we also recorded the postures of HoloLens using the built-in inertial sensors during the study, which corresponded to the participants’ head postures.

To quantitatively analyze participants’ movement behavior, we borrowed the concepts of macro-movements and micro-movements from Probst’s work [8], where she used the macro level and the micro level to categorize users’ movements in the office work environment. We developed a scheme (see Table 8.1) to code the participants’ movements from the recorded videos. Sit-stand transitions are the movements a participant sitting in the chair from a standing posture or standing up from a sitting posture. We regarded the posture of leaning to the wall or the table as standing when counting the sit-stand transitions. Body position/direction changes refer to the movements a participant moving her/his body or the chair (while sitting) to another position or direction. Macro-movements require large muscle groups working together. By contrast, micro-movements require less muscle effort, e.g., moving the head, legs, or torso postures. We particularly selected the movement of adjusting HoloLens because it is caused by the discomfort of wearing HoloLens. We have to separate it from other head movements, which includes neck-relaxing movements and intently moving the virtual screen in HoloLens. We did not count the movements of clicking the remote clicker or the air-tap gesture because they were only related to the participants’ speed of answering the questions and independent to the mode.

**Table 8.1: Macro-movements and micro-movements in our coding scheme.**

<table>
<thead>
<tr>
<th>Sub-groups</th>
<th>Macro-movements</th>
<th>Micro-movements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sit-stand transitions; Body position/direction changes.</td>
<td>Adjusting HoloLens; Head movements (excluding adjusting HoloLens); The rest (i.e., arms, legs, and torso movements).</td>
</tr>
</tbody>
</table>

One coder coded the videos when recording the study procedure. After the study, we invited one research assistant (objective coder) to randomly code a subset of the participants’ video data (Hallgren, 2012; Lazar, Feng, & Hochheiser, 2017) and then calculated the intra-class correlation (ICC) (Shrout & Fleiss, 1979) of the two coders. As we were only interested in the within-subject difference between the two experimental modes, we chose the consistency between the two coders to test the coding reliability. We selected two-hour video data for training.
the objective coder and five-hour video data for the formal coding. The ICCs were calculated for all the movement categories: sit-stand transitions (ICC = 1.00, p = 0.000), body position/direction changes (ICC = 0.79, p = 0.031), the overall macro-movements (ICC = 0.92, p = 0.005), adjusting HoloLens (ICC = 0.83, p = 0.020), head movements (ICC = 0.81, p = 0.026), micro-movements excluding the head-related ones (ICC = 0.08, p=0.444), and the overall micro-movements (ICC = -0.35, p=0.754). The coding consistency is good (ICC > 0.75) for the most categories. However, the coding for the micro-movements excluding the head-related ones showed poor reliability, which also caused poor reliability of the overall micro-movements. The large variance of the participants’ movement habits is the main difficulty of coding the micro-movements, e.g., rotating the chair slightly, shaking legs, swing body slightly. Therefore, we only focus on the measurements with good reliability in data analysis.

8.5 Statistical Analysis

As we discussed in 6.3.5, the conventional null-hypothesis significance tests provide little information when the result is not statistically significant – only the alternative hypothesis is tested (Dienes, 2014). Non-significant results might support a null hypothesis over the alternative, or the data are just insensitive. Therefore, we again adopt the Bayes factor (Kass & Raftery, 1995) in addition to the p-value (Greenland et al., 2016) and Cohen’s d (Fritz et al., 2012) to report and interpret the results. It compares the extent to what the samples support two hypotheses (e.g., equal or different) and allows more principled conclusions from small-n studies of novel techniques in the field of human-computer interaction (Kay et al., 2016). We use JASP\(^{21}\) (Version 0.9.2) for data analysis due to its ability of both the conventional null-hypothesis significance test and the corresponding Bayesian analysis. We still use the Bayes factor classification and the interpretation (Doorn et al., 2019) as shown in Figure 8.3.

![Figure 8.3: A graphical representation of a Bayes factor classification and the interpretation, adapted from (Doorn et al., 2019).](https://jasp-stats.org/)

Since we had prior information about the effect of the two modes on users’ behavior, we choose the two-sided alternative hypothesis \(H_1\) that the population mean of the difference is not equal to 0. Due to the same reason, we use the default Cauchy distribution \(r = 1/\sqrt{2}\) as the prior when we estimate the

\(^{21}\)https://jasp-stats.org/
8.6 Results

8.6.1 Macro-Movements

The t-test result shows no significant difference of macro-movement times between the two modes ($p = 0.109$, Cohen’s $d = 0.563$). The Bayes factor provides no evidence for both $H_0$ and $H_1$ ($BF_{10} = 0.993$, $M = 0.455$, $CI = [-0.137, 1.110]$). However, in the boxplot of the within-subject difference (Figure 8.4), we see a trend that the macro-movements of the tag-along mode are less than the spatial-mapping mode.

![Figure 8.4](image)

Figure 8.4: The boxplot of macro-movements in the spatial-mapping (SM) mode, and the tag-along (TA) mode. The boxplot of within-subject difference (TA-SM) is to show the effect intuitively. The circles in the boxplots refer to the means.

As macro-movements include sit-stand transitions and the movements of changing body direction/position, we decompose them into two categories for a deeper understanding (see Figure 8.5). The comparison of sit-stand transitions shows weak evidence for $H_0$ ($BF_{10} = 0.453$, $M = -0.243$, $CI = [-0.841, 0.310]$, $p = 0.359$, Cohen’s $d = -0.306$), while the result of body direction/position changes indicates moderate evidence in favor of $H_1$ ($BF_{10} = 3.324$, $M = 0.704$, $CI = [0.057, 1.488]$, $p = 0.022$, Cohen’s $d = 0.874$). Therefore, the participants changed their body direction/position when sitting or standing more often in the spatial-mapping mode than the tag-along mode. However, sit-stand transitions tended to be the same. This grouping analysis explains the comparison result of the total macro-

effect size. Following the JASP guidelines (Doorn et al., 2019), we also report the median ($M$) and the 95% credible interval ($CI$) of the effect size.

Regarding the result accuracy and reliability, we also checked the normality assumption, the robustness, and the error percentages of calculating the Bayes factors. Only one measurement violates the normality assumption, while all the other results showed low error percentage ($<=0.011\%$) and good robustness. For the one with a deviation from normality, we applied the Wilcoxon signed-rank test and calculated to matched pairs rank biserial correlation ($r$) (Kerby, 2014) to show the effect size.
movements: the sit-stand transitions data (weakly supporting H₀) compensates the moderate evidence for H₁ from the body direction/position changes data, resulting in no evidence for both H₀ and H₁ for the whole macro-movements.

Besides the number of sit-stand transitions, the duration of sitting/standing is also of interest. Because the data of standing duration violate the normality assumption, we use a Wilcoxon signed-rank test to compare the two modes. The result does not suggest a significant effect (p = 0.183, r = -0.709). However, from the boxplot (Figure 8.6) of the within-subject difference of the standing duration, we see most of the difference is positive except one outlier. In other words, all the participants but one stood more in the tag-along mode than the spatial-mapping mode.

8.6.2 Micro-Movements

Regarding the micro-movements, we investigate two categories: the movements of adjusting HoloLens (AH) and the rest head movements (see Figure 8.7). The AH involves several movements with hands and the head - caused by the discomfort of wearing HoloLens – showing weak evidence for H₁ (BF₁₀ = 1.660,
8.6 Results

M = 0.560, CI = [-0.044, 1.252], p = 0.054, Cohen’s d = 0.700). The rest head movements indicate strong evidence in favor of H₁ (BF₁₀ = 10.900, M = -0.980, CI = [-1.816, -0.154], p = 0.005, Cohen’s d = -1.168) with the other direction. In other words, the participants moved their heads much more frequently in the spatial-mapping mode than the tag-along modes, excluding the movements caused by adjusting HoloLens. Therefore, the comparison of head-related micro-movements shows different trends between the two modes to the overall micro-movements.

![Boxplot of head movements](image)

**Figure 8.7**: The boxplot of the numbers of adjusting HoloLens and head movements in the spatial-mapping (SM) mode, the tag-along (TA) mode, and the within-subject difference (TA-SM).

8.6.3 Head Direction

Besides movements, using right postures when sitting or standing is also critical to prevent musculoskeletal disorders (Grandjean & Hünting, 1977), especially the ones caused by new technologies (e.g., the “text neck” (Cuéllar & Lanman, 2017)). The pitch angle of the head corresponds to the pressure on the cervical vertebrae of the spine. The logged data of the HoloLens postures using the built-in sensors allow us to analyze the pitch angle of the participants quantitatively. We only include seven participants’ data because three participants’ data of the HoloLens posture were not complete due to technical issues. The Bayes factor shows very weak evidence in favor of H₁ (BF₁₀ = 1.092, M = -0.527, CI = [-1.335, 0.163], p = 0.114, Cohen’s d = -0.698). However, the box plot (Figure 8.8) shows a trend that the pitch angles in the tag-along mode are 6.2 degrees higher than the spatial-mapping mode on average. After checking each participant’s data, we found that the difference is mainly caused by the big changes of two participants (#5 and #6 in Figure 8.9). The medians of the pitch angle are 12.6 and 14.2 degrees in the two modes, which are very close to each other. It should be noticed that the participants changed their head pitch angles several times during the study (see Figure 8.7). However, the average values indicate that the participants spent more time using the head-up postures.
8.6.4 Workload

The Bayes factors of the workload assessment between the two modes show weak to moderate evidence in favor of \( H_0 \) (see Table 8.2 and Figure 8.10). This result indicates that the participants were under a similar workload during the two modes, as expected. Besides, the scores of the mental demand and the effort suggest that our task caused a moderate workload to the participants.

Table 8.2: The Bayes factors and the effect sizes (the medians and the credible intervals) of the workload assessment comparison between the two modes.

<table>
<thead>
<tr>
<th></th>
<th>Mental Demand</th>
<th>Physical Demand</th>
<th>Temporal Demand</th>
<th>Performance</th>
<th>Effort</th>
<th>Frustration</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF&lt;sub&gt;10&lt;/sub&gt;</td>
<td>0.414</td>
<td>0.318</td>
<td>0.653</td>
<td>0.411</td>
<td>0.743</td>
<td>0.655</td>
</tr>
<tr>
<td>M</td>
<td>0.204</td>
<td>0.063</td>
<td>0.345</td>
<td>-0.205</td>
<td>0.387</td>
<td>0.353</td>
</tr>
<tr>
<td>CI</td>
<td>[-0.350, 0.786]</td>
<td>[-0.480, 0.617]</td>
<td>[-0.223, 0.973]</td>
<td>[-0.796, 0.340]</td>
<td>[-0.191, 1.010]</td>
<td>[-0.228, 0.969]</td>
</tr>
</tbody>
</table>
8.7 Discussion

8.7.1 Explanations of Movements

The statistical result of the macro-movements comparison between the two modes moderately supports that the participants performed more movements of changing body position/direction in the spatial-mapping mode. The interview in our study provides us some hints to explain this result: when asked about the factors affecting them choosing postures and adjusting the virtual screen, the common answer of the participants was the physical and visual comfort. Because they could find a comfortable posture to do the task easier in the tag-along mode, they performed fewer movements to adjust their body positions/directions.

Regarding the sit-stand transitions, we see weak evidence in favor of the null hypothesis: no difference between the two modes. Based on the interview, we find that the reasons behind macro-movements are multifaceted. Here we list all the mentioned factors affecting the participants’ macro-movements:

1. The musculoskeletal tiredness/discomfort.
   “If getting tired when standing, I will sit down. When feeling uncomfortable using sitting postures, I stand up.”

2. The visual tiredness/discomfort.
   “I changed between sitting and standing to find a good posture to optimize visual comfort.”

3. The weight of HoloLens.
   “Sometimes I have to sit down and use my hand to hold HoloLens for reducing the weight on my head.”

   “I stood up because I thought it’s unhealthy to sit for a long time, but not I was uncomfortable when I sat there.”

5. Unconsciousness/habits.
“I think I just sat down after standing for a while because of my habit of sitting.”

The head movements excluding the ones of adjusting HoloLens in the tag-along mode were significantly more than the spatial-mapping mode with strong evidence. This result indicates that simplicity and freedom encourage head movements. We also found the reason from the interview: the participants needed to adjust postures to reduce the discomfort of wearing HoloLens and some static postures during the task; the tag-along mode provided an easier way to do so.

Based on the results of macro-movements and micro-movements (head movements), we can answer our research questions:

1. The higher degree of simplicity and freedom of moving the virtual screen leads to more head movements. However, the effect on macro-movements is complex. Weak evidence supports no effect on sit-stand transitions, while moderate evidence shows that the higher degree of simplicity and freedom lead to fewer movements of changing body position/direction.

2. The main factor affecting the participants’ movements during our task is the body discomfort including the musculoskeletal and the visual discomfort. The body discomfort might be caused by wearing HoloLens and the participants’ static postures. Some participants could remind themselves to stand up after sitting for health, but only occasionally.

Therefore, three factors are related to the participants’ movements during the study: the physical and visual discomfort of using HoloLens, the simplicity and freedom of moving the virtual screen, and the participants’ ability to remind themselves to move. Comparing to the traditional computer screen, HoloLens provides more movement and posture freedom, but less physical and visual comfort for screen-based tasks.

8.7.2 Reminders

The participants’ sit-stand transitions were around three times during the one-hour study session on average in both modes (see Figure 8.5), while the standing duration accounted for 20-30% of the one-hour session on average (see Figure 8.6). These numbers seem to indicate a healthy combination of sitting and standing durations. However, some participants still sat for a long time without moving. There were two participants even sitting all the time during the study (see Figure 8.11).

The score of the self-report habit strength index (5.47±0.65 out of 7) indicates that all the participants had moderate-to-strong sedentary habit strength. Figure 8.11 shows that the participants with sedentary work styles could still be sedentary even given the freedom of moving the virtual screen. Therefore, reminders for
prolonged sitting are also necessary when using AR-HMDs, just as the case when using computer screens.

![Figure 8.11](image1)

**Figure 8.11:** Each row represents a one-hour study session. The green part refers to standing; the red part refers to sitting; the white part means the participant stopped because of eye fatigue.

Besides the sitting duration, we also observed several bad postures when the participants wore HoloLens. The postures, which cause an increase of intradiscal pressure, might lead to a higher risk of musculoskeletal disorders over time (Grandjean & Hünting, 1977). Besides some bad postures we can usually see (e.g., slouching in the chair and crossing legs while sitting), we also observed some potentially harmful postures (see **Figure 8.12**) that we could hardly use with the traditional screen on the table. Therefore, posture monitoring and reminding are necessary, as we discussed in the section of related work. It could improve the existing AR-HMDs for health promotion by adding these features.

![Figure 8.12](image2)

**Figure 8.12:** Selected bad postures of the participants during our study.

All the participants agreed that a reminder would be helpful because they could easily forget to stand up and use bad postures when focusing on the task. Furthermore, three participants mentioned a prompt reminder might not work for them. They preferred alternative reminders to force them to change postures and move more, e.g., making the virtual screen tilt or blur to indicate bad postures and prolonged sitting.
8.7.3 Limitations and Future Work

This is an initial work to explore the potential impact of AR-HMDs on the movement behavior of office workers. There are several limitations to this work, which could also be research opportunities for future work. First, the user task in this study is simplistic. Real office work might involve more sophisticated user interaction, which should be considered in future work, e.g., browsing webpages or navigating maps.

Second, the wearing comfort of HoloLens might limit the generalization of this work. Due to the disadvantages of weight and the display quality (field of view and resolution) in HoloLens, all participants felt uncomfortable after each one-hour study session. Future work could use other AR-HMDs with a higher degree of wearing comfort. Furthermore, using other AR systems (e.g., RoomAlive (Jones et al., 2014)) to study users’ movement behavior avoiding the effect of discomfort of wearing HMDs also deserves future studies.

Third, we only designed one virtual screen in our study, which did not make full use of the spatial-awareness feature of HoloLens. It would be interesting to study the impact of multi-screen on users’ movements.

Lastly, the sample size is relatively small. Despite the moderate to strong evidence in the Bayesian analysis, some evidence is very weak, e.g., the sit-stand transitions and the HoloLens pitch angle. The weak evidence might be improved by using larger sample size.

8.8 Conclusion

Through the exploratory study, we have several interesting findings: (1) moderate evidence supports that body direction/position changes were more in the spatial-mapping mode than the tag-along mode; (2) weak evidence supports that sit-stand transitions and standing durations had no difference between the two modes; (3) the participants adjusted HoloLens more times in the spatial-mapping mode with weak evidence; (4) the participants moved the head more times (excluding the ones related to adjusting HoloLens) in the tag-along mode with strong evidence. The simplicity and freedom of moving the virtual screen encouraged movement behavior, while the discomfort of wearing HoloLens could also make users move to adjust it. All the participants preferred the tag-along mode because they felt it more comfortable, but we also observed bad postures as the side effect of the high simplicity and freedom of moving the virtual screen.

To sum up, this work made the following contributions to the research field of multimedia and human-computer interaction: (1) it is the first work to investigate the potential impact of the augmented-reality head-mounted display (AR-HMD) on users’ movement behavior in screen-based office work; (2) inspired by related work, we analyzed the movement behavior through the lenses of macro-
movements and micro-movements with categorizing the movements to sub-groups; (3) we used Bayesian method, in addition to the null-hypothesis significant test, to analyze and report the study results; (4) the study results confirmed the effect of the freedom and simplicity (EoFS) of moving the virtual screen in HoloLens on users’ movement behavior; (5) besides EoFS, we also found that the discomfort of using HoloLens could partially cause users’ movement behavior; (6) based on the findings and the limitations of this work, we provide four research opportunities for future work.

Through this study, we want to show the necessity of studying the ergonomics of AR-HMDs, especially their impact on users’ movement behavior. Besides the exciting user experience brought by the new technology, we should also consider the health perspectives when designing applications using AR-HMDs.
9 Health Behavior Change in HCI: Trends, Patterns, and Opportunities

不识庐山真面目，只缘身在此山中。

One cannot see the forest for the trees.

- SU Shi (1037 ~ 1101)
(A Chinese writer, poet, painter, calligrapher, pharmacologist, gastronome, and a statesman of the Song dynasty)


Parts of this chapter appear in the above publication. The responsibilities for this joint publication were divided as follows: I formulated the research question, designed and conducted the study, analyzed the study data, and spearheaded the writing. Harald Reiterer supervised the work and reviewed this paper.
9.1 Abstract

Unhealthy lifestyles could cause many chronic diseases, which bring patients and their families much burden. Research has shown the potential of digital technologies for supporting health behavior change to help us prevent these chronic diseases. The HCI community has contributed to the research on health behavior change for more than a decade. In this chapter, we aim to explore the research trends and patterns of health behavior change in HCI. Our systematic review showed that physical activity drew much more attention than other behaviors. Most of the participants in the reviewed studies were adults, while children and the elderly were much less addressed. Also, we found there is a lack of standardized approaches to evaluating the user experience of interventions for health behavior change in HCI. Based on the reviewed studies, we provide suggestions and research opportunities on six topics, e.g., game integration, social support, and relevant AI application.

9.2 Introduction

The research on digital technologies to support health behavior change is no doubt a vital task for the Human-Computer Interaction (HCI) community. Only in the proceedings of the ACM CHI conference until 2018, we found 310 papers mentioning “behavior change.” However, it seems that the interest in health behavior change from the HCI community began to decrease recently. We see this trend by searching and screening the related papers from the ACM digital library. The amounts of the related papers from both the CHI conference and the UbiComp conference have seen the decrease since 2016, and the corresponding paper amount in CHI 2018 has fallen back to the level in 2014 (see Figure 9.1). To get an insight into this phenomenon, we conducted a systematic review of the papers about health behavior change from the HCI community. We extracted information from the perspectives of target behaviors, target user groups, the used behavioral theories, the deployed behavior change strategies, intervention characteristics, and evaluation methods.

The remainder of this chapter is organized as follows: The next section introduces behavioral theories, behavior change strategies, and behavioral intervention characteristics as the apparatus of our review. In Section 9.4, we show our methods to search, select, and code the studies. Section 9.5 reports our findings on research trends and patterns of health behavior change in HCI. Based on our reviewed studies, in Section 9.6, we provide suggestions and opportunities for the future research in six topics. Finally, we show the limitations of our work and conclude the chapter.
Figure 9.1: The paper frequency distribution in the field of health behavior change in the HCI community. Note that we extracted the original data from the ACM digital library on August 23, 2018, when some conferences for this year have not taken place. The majority of the papers are from conference proceedings, while a small part of them are from journals (e.g., Personal and Ubiquitous Computing or PUC in the figure).

9.3 Background

9.3.1 Behavioral Theories

Behavioral theories refer to the social-psychological theories of behavior change, which explain and predict human behavior. Glanz et al. (Glanz K, Rimer BK, 2008) listed the most frequently used behavioral theories published before 2010: the Social Cognitive Theory (SCT) (Bandura, 1977a), the Transtheoretical Model of Change (TTM) (Prochaska & Velicer, 1997), the Health Belief Model (HBM) (Rosenstock et al., 1988), and the Theory of Planned Behavior (TPB) (Ajzen, 1991). As explained by Sutton (Sutton, 2002), each of the behavioral theories specifies a small number of cognitive and affective factors as the proximal determinants of behavior.

In a CHI paper in 2013, Hekler and colleagues illustrated the (dis)advantages of behavioral theories and how HCI researchers can use and contribute to behavioral theories (Hekler et al., 2013). In summary, behavioral theories can help inform design, guide evaluation strategies, and select target users. Also, HCI researchers have the change to improve behavioral theories by improving measurement, enhancing early-stage theory fidelity testing, and supporting and using big data and A/B testing. Following this implication, we will reveal how HCI research engaged with behavioral theories in the real world.
9.3.2 Behavior Change Techniques (Strategies)

Behavior change techniques (BCTs) are defined as observable, replicable, and irreducible components of an intervention designed to change behavior (Abraham & Michie, 2008; Michie et al., 2013), e.g., self-monitoring of behavior and goal setting. Abraham and Michie published the taxonomy containing 93 BCTs in 16 groups in 2013, called Behavior Change Technique Taxonomy (v1) (Michie et al., 2013). The BCT taxonomy has been used for informing intervention development (Sarah A. Mummah et al., 2016; Sarah Ann Mummah, Mathur, et al., 2016) and identifying the effective ingredients in intervention studies for health behavior change (Dombrowski et al., 2012; Gardner et al., 2016; Michie et al., 2009; Olander et al., 2013) and products (i.e., health-oriented apps (Conroy et al., 2014; Crane et al., 2015; Direito et al., 2014; Middelweerd et al., 2014) and wearables (Lyons et al., 2014)). The word cloud in Figure 9.2 shows the relative use frequencies of BCTs used in 405 studies. In the HCI community, however, BCT taxonomy is not used as widely as the model of Persuasive Technology (Fogg, 2003) or Persuasive System Design (PSD) (Oinas-kukkonen & Harjumaa, 2009). The model of PSD includes 28 principles in four categories, namely primary task support, dialogue support, system credibility support, and social support.

Figure 9.2: The word cloud of behavior change strategies coded from the papers listed on the official website of BCT taxonomy (N=405). The bigger the font is, the more frequently the strategy was used.

In comparison with PSD principles, BCT taxonomy provides a more comprehensive pool of strategies for behavior change interventions. Even though, BCT taxonomy does not cover all the strategies we have found in related studies. Therefore, we add another five strategies to BCT taxonomy for our coding, which include social cooperation, social competition, social recognition, virtual reward, and game integration. The former three are derived from PSD principles. By game integration, we mean both exergames (Mueller, Gibbs, & Vetere, 2009) and gamification (Miller, Cafazzo, & Seto, 2016).
9.3.3 The Behavioral Intervention Characteristics

The behavior change strategies are only about “what” but not “how” of the intervention. In 2014, Mohr and colleagues proposed the behavioral intervention technology (BIT) model to support the translation from behavior change intervention aims into an implementable treatment model (Mohr et al., 2014). Inspired by the BIT model, we include the concepts of intervention characteristics and intervention workflow in our coding, which can help us analyze how interventions are delivered. The characteristics include the medium, the visualization method (related to aesthetics), and the social support type in our coding. We further elaborate the details in the following section.

9.3.4 The Holistic Framework

In Chapter 3, we proposed a holistic framework to guide the design and report of health behavior change interventions (Y. Wang, Fadhil, Lange, & Reiterer, 2019). This framework integrates the three mentioned aspects in this section. By following this framework, we aim to provide the most comprehensive review of health behavior change in HCI. We emphasize comprehensiveness and consistency in reviewing health behavior change because health behavior change studies are always complex processes and affected by many aspects in field studies.

9.3.5 Related Work

In a highly related work, Orji and Moffatt (Orji & Moffatt, 2018) reviewed how persuasive technology was used for health and wellness in 85 related papers. They coded the reviewed studies from 11 perspectives: targeted (health) domain, technology, duration of evaluation, behavior theories, motivational strategies, evaluation method, targeted age group, number of participants, study country, targeted behavioral or psychological outcome, and findings/results. However, the coding of motivational strategies did not follow any existing taxonomy of persuasive technology (e.g., PSD principles) or behavior change techniques (e.g., BCT taxonomy). Thus the definitions of these strategies can be vague and imprecise for readers. We use a taxonomy integrating BCTs and PSD principles to code and analyze the adopted digital health strategies. Using the holistic framework illustrated in Chapter 3 (Y. Wang, Fadhil, Lange, et al., 2019) can help to achieve a more comprehensive review of the related studies.

Since existing systematic review on the effectiveness of digital health interventions have pointed out that there are not enough high-quality studies to draw powerful conclusion on effectiveness - e.g., eating behavior change (McCarroll, Eyles, & Ni Mhurchu, 2017) and sedentary behavior change at work (Y. Wang, Wu, et al., 2018) - we put our effort on revealing the patterns and trends of the existing empirical studies. We focus on multiple health behaviors instead
of a specific one because we want to find out the patterns in different target behaviors.

9.4 Methods

As our initial aim is to find the research trends and patterns of health behavior change in the HCI community, we only used ACM digital library as our search repository, which covers most of HCI conference proceedings (e.g., CHI and UbiComp). For the searching, we considered the spelling versions of behavior/behaviour, similar expressions of behavior/behavioral change, persuasive technology, and the names of targeted behavior (e.g., physical activity and alcohol). We also excluded the papers focusing on sustainability, since they are out of the scope of this chapter. The search was conducted on 23rd Aug, 2018, and the query syntax we used in ACM digital library is shown in Table 9.1.

<table>
<thead>
<tr>
<th>Table 9.1: The query syntax used in ACM digital library.</th>
</tr>
</thead>
<tbody>
<tr>
<td>keywords.author.keyword:(+behavior +change -sustainability) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(+behavioral +change -sustainability) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(+behaviour +change -sustainability) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(+behavioural +change -sustainability) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(+persuasive +technology -sustainability) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(+physical +activity) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(diet dietary) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(+sexual +health) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(smoking) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(sleeping) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(sedentary sitting) OR</td>
</tr>
<tr>
<td>keywords.author.keyword:(alcohol)</td>
</tr>
</tbody>
</table>

The four-phase flow diagram of PRISMA (Liberati et al., 2009) was used to illustrate the study selection process (see Figure 9.3). A total of 1070 papers were identified out of 530,358 records in ACM digital library. The first coder screened the records from the paper title. 354 records were screened out because of duplication (N=15), not being relevant to health behavior change (N=337), or being the workshop introduction (N=2). Afterward, the first coder reviewed the abstracts (or full paper if necessary) for the rest of the papers (N=776) and labeled the papers by the research method (see Table 9.2 for details) and target behavior. We further excluded the papers of duplication (N=10), not about health behavior change (N=72), workshop introductions (N=2), panel introductions (N=3), Ph.D. colloquia (N=20), courses & tutorials & talk introductions (N=4), and not in English (N=2). Finally, we obtained 648 papers falling into the listed types in Table 9.2. The paper list can be found in the supplementary material.
Chapter 9: Health Behavior Change in HCI: Trends, Patterns, and Opportunities

Figure 9.3: The workflow of screening and selecting papers.

Table 9.2: The paper types used in our coding.

<table>
<thead>
<tr>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compared intervention study</td>
<td>It includes at least one user intervention session with at least two compared conditions.</td>
</tr>
<tr>
<td>Exploration study</td>
<td>It includes at least one user intervention session for behavioral factors exploration without compared conditions.</td>
</tr>
<tr>
<td>Test study</td>
<td>It is a feasibility test or pilot study without enough measures of users’ behavior or outcome.</td>
</tr>
<tr>
<td>Design</td>
<td>It is about designing systems or methods to support health behavior change without any user intervention.</td>
</tr>
<tr>
<td>Interview</td>
<td>It includes only user interviews without any user intervention.</td>
</tr>
<tr>
<td>Survey</td>
<td>It includes only surveys based on questionnaires without any intervention to users.</td>
</tr>
<tr>
<td>Data mining</td>
<td>It is about systems/algorithms to detect, recognize, classify, or predict human behavior or behavioral factors for health behavior change.</td>
</tr>
<tr>
<td>Review</td>
<td>It overviews or reviews previous work.</td>
</tr>
<tr>
<td>Framework &amp; Theory</td>
<td>It proposes frameworks or theories for the research on health behavior change.</td>
</tr>
<tr>
<td>Viewpoint</td>
<td>It provides viewpoints, guidelines, or implications for the research on health behavior change.</td>
</tr>
<tr>
<td>Concept</td>
<td>It includes only concepts of systems or methods for health behavior change.</td>
</tr>
</tbody>
</table>
From the 648 papers in the phase of eligibility, we selected 72 papers that include compared intervention studies (73 studies in total). Afterward, two coders coded these full papers separately, and the differences were resolved by discussion. The coding schema is shown in Table 9.3.

9.5 Results and Findings

In this section, we report the results and findings of the systematic review based on the methods introduced in the previews section. We firstly show the trends of the paper amount in the perspectives of the target behavior and the paper type (see Table 9.2) over the research history of health behavior change in HCI. Then we select “compared intervention studies” (N=75) and analyze the research patterns in the views of measurements, user experience evaluation, the target behavior, the target user group, the application of behavioral theories, the use of behavior change strategies, and intervention characteristics.

<table>
<thead>
<tr>
<th>No.</th>
<th>Item</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Target behavior</td>
<td>Physical activity, diet, etc.</td>
</tr>
<tr>
<td>2</td>
<td>Target user group</td>
<td>Adults, children, etc.</td>
</tr>
<tr>
<td>3</td>
<td>Behavioral theory</td>
<td>TTM, goal-setting theory, etc.</td>
</tr>
<tr>
<td>4</td>
<td>Behavioral theory use type</td>
<td>Informing design; guide evaluation strategies; selecting target users.</td>
</tr>
<tr>
<td>5</td>
<td>Behavior change strategy</td>
<td>BCT Taxonomy (v1) + (cooperation, competition, Recognition, virtual reward, and game integration).</td>
</tr>
<tr>
<td>6</td>
<td>Measurement</td>
<td>User experience (quantitative); user experience (qualitative); target behavior; user interaction (i.e., use frequency, use duration); behavioral factors (e.g., constructs from behavioral theories).</td>
</tr>
<tr>
<td>7</td>
<td>User experience Instrument</td>
<td>SUS (Brooke, 2013), AttrakDiff22, etc. (Coding only when user experience is quantitatively measured.)</td>
</tr>
<tr>
<td>8</td>
<td>Intervention workflow</td>
<td>Time-based; task-based; event-based. (Coding only when prompts/cues are used as a behavior change strategy.)</td>
</tr>
<tr>
<td>9</td>
<td>Intervention Characteristic</td>
<td>See Table 9.4 for details.</td>
</tr>
</tbody>
</table>

Table 9.3: The coding items and explanations.

22 http://attrakdiff.de
### 9.5.1 Trends of Target Behaviors and Paper Types

We analyze the trends based on the papers after the title and abstract screening (N=648). Of the 48 target behaviors we found, five behaviors (i.e., physical activity, sleep, diet, smoking, and sedentary behavior) account for 73% in all the papers. As shown in Figure 9.4, physical activity remains the most popular target behavior over time and the corresponding papers keep growing in the last six years. The number of the papers targeting sedentary behavior also peaked in 2017. The paper amounts for sleep, diet, and others decreased in 2017 after 2-4 years’ increase. From the perspective of the target behavior, the decrease of papers about sleep, dietary behavior, and other behaviors except the ones listed in Figure 9.4 caused the drop in the overall paper count in 2017.

![Figure 9.4](image)

**Figure 9.4**: The target behaviors in the reviewed papers over time (N=648).

**Figure 9.5** illustrates the change of the paper amount regarding the paper type over time. Most (55%) of the papers contain designing new intervention systems or methods for health behavior change. However, only about 25% of the developed systems or methods were evaluated by the intervention study with compared conditions. The “data mining” papers saw a drop in 2017, while the “survey” papers and the “interview” papers have been rising in the last three
9.5 Results and Findings

years. The drop in “data mining” and “design” papers mainly contributed to the overall decrease in 2017.

![Figure 9.5: The paper category in the reviewed papers over time (N=648).](image)

9.5.2 Patterns in the Selected Intervention Studies

9.5.2.1 Measurement & User Experience Evaluation

Differing from the target behavior as shown in Figure 9.4, the user interaction means the objective measure of how the users use the intervention system (e.g., the use frequency). In comparison with the user experience, the behavioral factors refer to the constructs (e.g., self-efficacy) influencing the behavior change process. The target behavior of users was measured in most of the studies (93%), as shown in Figure 9.6. More than half of the studies (59%) evaluated user experience quantitatively or qualitatively. The user interaction with the intervention system was measured in 34% of the studies. Only 20% of the studies accessed users’ behavioral factors, which is related to the usage of behavioral theories.

![Figure 9.6: The percentages for the used measurements in the reviewed studies.](image)
Although about 59% of the studies accessed the user experience, only 32% (24/75) of them evaluated the user experience with quantitative measurements. The system usability scale (SUS) was used in four studies (H. Du, Youngblood, & Pirolli, 2014; J. Du, Wang, de Baets, & Markopoulos, 2017; Macvean & Robertson, 2013; Zuckerman & Gal-Oz, 2014), while the NASA-TLX (Komninos, Dunlop, Rowe, Hewitt, & Coull, 2015) and the AttrakDiff (Dantzig, Geleijnse, & Halteren, 2013) were used only once. The studies with game integration were more likely to measure users’ perceived enjoyment (e.g., (Berkovsky, Coombe, Freyne, Bhandari, & Baghaei, 2010; Hagen, Chorianopoulos, Wang, Jaccheri, & Weie, 2016; Malinverni, Silva, & Parés, 2012)). One study (Yun & Arriaga, 2016) was conducted within a clinical trial, which used the Patient Reaction Assessment (PRA) questionnaire to measure users’ experience of the intervention. We did not find any specific scale to evaluate the user experience of interventions for health behavior change.

9.5.2.2 Target User Group & Behavior

As shown in Figure 9.7, the target user group in most studies was the adult (68%), while almost half of these studies used college students and staff as the participants. Children, as the target user group, accounted for 15% in all studies. There is only one study targeting teenagers, while one study focused on young adults. From other aspects of the user group: five studies aimed at patients, three studies focused on the female, and one study only considered athletes. The reviewed studies are very unbalanced regarding the target user group.

![Figure 9.7: The distribution of the target user group.](image)

9.5.2.3 Behavioral Theories

Among the 75 selected intervention studies, 32 studies (43%) explicitly described the use of behavioral theories. The transtheoretical model (TTM) was the most frequently used theory, which was adopted in eight papers. This result is in line with another systematic review by Orji and Moffatt (Orji & Moffatt, 2018). The other behavioral theories adopted in the reviewed studies are listed in Figure 9.8.
9.5 Results and Findings

Figure 9.8: The distribution of the used behavioral theories. SRT refers to the self-regulation theory; HBM refers to the health belief model; IBM refers to the integrated behavioral model. The other acronyms can be found in the following content.

Figure 9.9 illustrates how behavioral theories were used in the reviewed studies. The TTM was mainly used to select target users (Consolvo, Everitt, Smith, & Landay, 2006; Consolvo et al., 2008; Gouveia et al., 2016; Kanaoka & Mutlu, 2015; Komninos et al., 2015; Lim, Shick, Harrison, & Hudson, 2011), as illustrated in (Hekler et al., 2013). In the case of using the TTM to inform the intervention design, different strategies were delivered according to the user’s stage of change (Graham et al., 2006; Kamphorst, Klein, & Wissen, 2014). The self-efficacy theory (SET) (Bandura & Adams, 1977), the theory of planned behavior (TPB), the self-determination theory (SDT) (Deci & Ryan, 2000), the Fogg’s behavior model (Fogg, 2009a), and the goal-setting theory (GST) (Locke & Latham, 2002) largely contributed to informing the intervention design.

Regarding the studies that did not utilize behavioral theories, we found that 29% (12 studies) focused on exergame and gamification (e.g., (Schäfer, Bachner, Pretscher, Groh, & Demetriou, 2018)), while 21% (9 studies) targeted children or teenagers (e.g., (Toscos, Faber, An, & Gandhi, 2006)). Behavioral theories might be useless in the case of the short game session (e.g., exergame). However, users’ adoption and engagement with health orientated game design could also be explained by behavioral theories. The work from Macvean and Robertson (Macvean & Robertson, 2013) indicated that children’s motivation of playing exergame would decrease over time and self-efficacy theory can predict and interpret the longitudinal physical activity patterns of children’s behavior change as well.
Figure 9.9: Behavioral theories and the ways that they were used in the reviewed studies. I – informing design, G – guide evaluation strategies, S – selecting target users. Note that, in this alluvial graph, the relative sizes of the bars for each behavioral theory do not exactly represent their use frequency. In one study, a behavioral theory can be used for both informing design and guiding evaluation strategies.

9.5.2.4 Behavior Change Strategies

Among the reviewed studies, the most used behavior change strategies are self-monitoring of behavior, goal-setting (behavior), feedback on behavior, prompts/cues, and game integration (see Figure 9.10). In Section 9.3.2 we have shown the cloud map of behavior change strategies coded from the papers listed on the website of BCT taxonomy. Those papers are mainly from journals of public health, behavioral science, and healthcare (e.g., BMC public health23 and JMIR24). In comparison with our reviewed papers, the researchers of those papers are more likely to use goal-setting (behavior), action planning, problem-solving, instruction on how to perform a behavior, and information about health consequences. This indicates the different use patterns of behavior change strategies between the HCI community and the community of public health, behavioral science, and healthcare.

---

23 [https://bmcpublichealth.biomedcentral.com/](https://bmcpublichealth.biomedcentral.com/)
9.5 Results and Findings

Figure 9.10: The word cloud of behavior change strategies found in our reviewed intervention studies (N=75). The bigger the font is, the more frequently the strategy was used.

We have shown the distribution of the target behavior in the papers of all the selected types (N=648) in Figure 9.4, where the papers targeting physical activity are much more than other types. Among the 75 full coded studies, the target behaviors are also unbalanced in quantity, and physical activity is still the most addressed behavior. Figure 9.11 illustrates the interaction of the top-4 target behaviors and behavior change strategies (in-group) in the reviewed studies. The alluvial graph indicates that a variety of behavior change strategies was used for all the target behaviors. One interesting finding is that almost all of the studies using game integration were designed for physical activity.

Figure 9.11: The top-4 target behaviors and the corresponding behavior change strategies groups in the reviewed studies. Note that, in this alluvial graph, the relative sizes of the bars for each target behavior do not exactly represent their frequency. Given the target behavior in one study, several strategies might be used.
9.5.2.5 Intervention Workflow

We found 16 out of 75 studies involving reminders (i.e., prompts/cue), including 9 event-based reminders (Berque et al., 2011; Birk & Mandryk, 2018; Caraban, Karapanos, Campos, & Gonçalves, 2018; Dantzig et al., 2013; J. Du et al., 2017; J. Lee et al., 2017; Shin et al., 2016; Spelmezan, Jacobs, Hilgers, & Borchers, 2009), and 7 time-based reminders (Bentley & Tollmar, 2013; Freyne, Brindal, Hendrie, Berkovsky, & Coombe, 2012; Kocielnik & Hsieh, 2017; Munson, Krupka, Richardson, & Resnick, 2015; Pater, Owens, Farmer, Mynatt, & Fain, 2017; Stawarz, Cox, & Blandford, 2015; Yun & Arriaga, 2016). We did not find any task-based intervention workflow, according to the definition in the BIT model (i.e., the release of intervention elements are based on the user’s completion of prescribed intervention tasks (Mohr et al., 2014)). Among the studies where the intervention system did not provide any scheduled reminders or prompts, we found a group of studies using always-on glanceable cues (Bauer et al., 2012; Consolvo et al., 2008; Gouveia et al., 2016; Rogers, Hazlewood, Marshall, Dalton, & Hertrich, 2010) to nudge users. For example, Gouveia and colleagues (Gouveia et al., 2016) designed watch faces of the smartwatch to provide real-time feedback about the user’s physical activity.

9.5.3 Characteristics of the Selected Intervention Studies

9.5.3.1 Media

The media determine the information channel of the intervention. The mobile phone (including the smartphone and the functional phone) was used in most of the studies (44/75). The mobile phone, especially the smartphone, has become indispensable in our daily life. Therefore, the high adoption rate of the mobile phone is not surprising. The rest of the adopted media in the studies are listed in Figure 9.12. The web means that the study did not restrict users to use a mobile device or a PC. Except for the common devices (e.g., PC, the mobile phone, the fitness tracker), some new devices were created to solve specific problems. For example, the wearables for monitoring sitting posture (J. Du et al., 2017) and augmented slider for supporting children’s learning process (Malinverni et al., 2012).
9.5 Results and Findings

Figure 9.12: The media used in the reviewed studies. The unknown refers to the studies that did not explicitly mention the media. The wearable means the ones users can wear on clothes or shoes, rather than fitness trackers and smartwatches.

9.5.3.2 Visualization

The visualization means how the intervention is presented to the users via the software interface. As shown in Figure 9.13, the plain text (e.g., SMS and notification), the progress bar, and the gamification interface were the most popular visualization methods. The others include the virtual agent, the timeline, the leaderboard, the reward sheet, the icon, the cartoon figure, the Emoji, and so on.

Figure 9.13: The word map of the visualization methods used in the reviewed studies. The bigger the font it, the more frequently the visualization method was used.

We found 21 studies that provided the function of social support. The types of social support appeared in these studies are listed in Table 9.5. Some of the types are included in the BCT taxonomy (Michie et al., 2013) (e.g., social comparison and social incentive) and PSD principles (Oinas-kukkonen & Harjumaa, 2009) (e.g., social cooperation and social competition). However, our social support types and the explanations might be different from the definitions in the BCT taxonomy and PSD principles. Instead, we derived these social support types by analyzing the intervention descriptions in the reviewed studies.
Table 9.5: The social support types.

<table>
<thead>
<tr>
<th>No.</th>
<th>Social Support Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Social commitment</td>
<td>It allows a user to make commitments within the intervention platform (Munson et al., 2015).</td>
</tr>
<tr>
<td>2</td>
<td>Social sharing</td>
<td>It allows users to see each other’s status, without aiming to compare with each other within the intervention platform (H. Du et al., 2014; Kim, Kientz, Patel, &amp; Abowd, 2008).</td>
</tr>
<tr>
<td>3</td>
<td>Social comparison</td>
<td>It allows users to compare with each other within the intervention platform (Chiu et al., 2009; Consolvo et al., 2006; Fialho et al., 2009).</td>
</tr>
<tr>
<td>4</td>
<td>Social competition</td>
<td>It allows users to compete with each other within the intervention platform (Foster, Linehan, Kirman, Lawson, &amp; James, 2010; Hagen et al., 2016; Nakanishi &amp; Kitamura, 2016; Zuckerman &amp; Gal-Oz, 2014).</td>
</tr>
<tr>
<td>5</td>
<td>Social communication</td>
<td>It allows users to communicate with each other within the intervention platform (Boratto, Carta, Mulas, &amp; Pilloni, 2017; Consolvo et al., 2006).</td>
</tr>
<tr>
<td>6</td>
<td>Social incentive</td>
<td>It allows users to encourage each other within the intervention platform (Chiu et al., 2009; H. Du et al., 2014; Katule, Rivett, &amp; Densmore, 2016; Shin et al., 2016).</td>
</tr>
<tr>
<td>7</td>
<td>Social interaction</td>
<td>It allows users to interact with each other within the intervention tool face to face (Bekker, Sturm, &amp; Eggen, 2010; Ludden &amp; Meekhof, 2016).</td>
</tr>
<tr>
<td>8</td>
<td>Social monitoring</td>
<td>It allows other users to monitor a user’s behavior, but not vice versa (Pater et al., 2017; Shin et al., 2016; Yuda, Kurata, Yoshida, &amp; Hayano, 2017).</td>
</tr>
<tr>
<td>9</td>
<td>Social recognition</td>
<td>It allows users to recognize their peers in public (Kitagawa, 2015; Rogers et al., 2010).</td>
</tr>
</tbody>
</table>

### 9.6 Discussion

We have reported our findings of the research trends and patterns of the research on health behavior change in the HCI community. These findings indicate several shortcomings and problems to be addressed in this research field:

1. The target behaviors mainly fell into physical activity, while some critical behaviors (e.g., sedentary behavior (Y. Wang, Wu, et al., 2018) and stress management (Sano et al., 2017)) were much less addressed.
2. The target users or study participants were mostly adults who were mostly college students and staff. We believe that more attention on the elderly is required in the aging society.
3. There is no standardized approaches or instruments for evaluating the user experience of intervention systems for health behavior change.
4. There is no standard to report the intervention study for health behavior change. The method we used to review the related papers provide an approach to reporting the relevant study. The study aspects that we suggest to report are shown in Table 9.3 and Table 9.4.
Following, we select six topics inspired by the reviewed studies to provide suggestions and opportunities for future studies. The first three topics are about users: considering the user’s behavior priority, categorizing target users from different perspectives, and leveraging users’ power of creativity and engagement. The other topics include longitudinal studies with game integration, cautions for socialization, and the applications of AI in health behavior change.

9.6.1 Users’ Behavior Priority

Although behavioral theories could be beneficial for the research on health behavior change in HCI, they are not without limitations. One of the limitations is that behavioral theories can explain only 20-30% of the total variance in a given health behavior (Hekler et al., 2013). From the reviewed papers, we noticed one factor that could collaboratively explain health behavior change. In the study by Rodgers and colleagues (Rodgers, Maloney, Ploderer, & Brereton, 2016), they found that college students consciously prioritize academic success over a healthy sleeping pattern. This finding indicates the fact that an individual’s daily life is filled with various tasks and behaviors (e.g., academic success and healthy sleeping pattern), instead of only the target behavior of a given intervention study. People can fail to adhere to an action plan just because they need to do other actions with higher priority in their limited time. Systems that can support users to schedule their daily activity and fit the target behavior into their routine could be a solution to the difficulty caused by priority.

Therefore, we suggest that future intervention studies should consider the behavior priority of participants. For example, the efficacy of sedentary behavior interventions could relate to users’ work priority. Intervention designers should check if there are critical tasks or dues hindering users’ enactment of their plans to reduce their sedentary behavior.

9.6.2 Categorizing Target Users

An intervention may be valid only for a specific group of audience, and the lack of specification of users could hide the effectiveness in study results. For example, Lacroix and colleagues (Lacroix, Saini, & Holmes, 2008) found that positive linear relationship between goal difficulty and users’ performance only existed for inactive people, but not for active people. Therefore, categorizing the target users in meaningful ways could lead to a better understanding of the intervention efficacy. In Figure 9.9, we have shown that the transtheoretical model was often used to group participants into different stages of change and select the target users. Besides the stage of change, researchers have found other methods and perspectives to categorize target users. E.g., Wiafe and colleagues (Wiafe, Nakata, & Gulliver, 2014) proposed a model to classify users based on their current behavior, attitude, and levels of cognitive dissonance.
Users’ personality is also a potential way to categorize target users, e.g., the well-known Big Five personality traits (de Vries, Truong, Zaga, Li, & Evers, 2017). As many health intervention studies have used gamification, the research on the personality of users (players) in games has drawn more attention (Nacke, Bateman, & Mandryk, 2014). Orji et al. (Orji, Nacke, & Di Marco, 2017) examined how different personalities respond to various persuasive strategies that are used in persuasive health games and gamified systems. Another study (Orji, Mandryk, & Vassileva, 2017) showed that tailoring the game design to players’ personality type improved the effectiveness of the games in promoting positive attitudes, intention to change behavior, and self-efficacy.

Our second suggestion is that future intervention could categorize target users not only based on the stage of change but also other factors (e.g., personality).

9.6.3 Leveraging Users’ Power

Researchers have started to explore and leverage the users’ creativity in health behavior change. Lee and colleagues (J. Lee et al., 2017) deployed a self-experiment design to support behavior change for improving participants’ sleep quality. In another study (Pater et al., 2017), a participatory design session was used after an intervention session for medication management among the elderly. Both of the studies proved the efficacy of user participation in the intervention design process.

In the study of Birk and Mandryk (Birk & Mandryk, 2018), a group of users was asked to customize their avatars to interact in a breathing exercise program. Compared to the control group with randomly assigned avatars, the customization group saw significantly less attrition and more sustained engagement through the 3-week study. In this study, customizing an avatar with its appearance and attributes required a minimum of 4-minute work of a participant. The effect of users’ participation can be explained not only from the perspective of customization but also from the view of IKEA effect (Norton, Mochon, & Ariely, 2012; Y. Wang, Pfeil, & Reiterer, 2016). The involvement of users’ effort in a product can increase their evaluation of the product.

Based on the evidence, we suggest that future intervention designers should take advantage of users’ participation and further explore the effect and user experience of participation.

9.6.4 Longitudinal Study with Game Integration

By game integration, we mean both exergames (i.e., interactive games that require players to invest significant physical effort as part of the gameplay (Mueller et al., 2009)) and gamification (i.e., implementing the most common and enjoyable mechanics of video games in non-video game contexts) (Miller et al., 2016). We extracted 15 studies with game integration and found that the study duration tended to be short, as shown in Figure 9.14. Macvean and Robertson
(Macvean & Robertson, 2013) studied children’s physical activity patterns when using an exercise game on smartphones over seven weeks, which is the longest study on gamification among the 15 studies. Seven studies only reported their evaluation for one game session, which we counted as one day in Figure 9.14.

![Figure 9.14: The study duration of the studies with game integration.](image)

The distribution of the reviewed study durations is shown in Figure 9.15, where we can see the number of studies with the duration less than one day is 13. More than half of the short-term studies are about game integration. Therefore, more longitudinal studies in gamification are required, because the goal of health behavior change is to help users maintain a healthy lifestyle in the long term.

![Figure 9.15: The distribution of the reviewed study durations.](image)

9.6.5 Cautions for Socialization

Since every individual is part of the social network and mobile technologies keep changing our communication in the social network, it is vital to investigate how socialization can benefit health behavior change. Socialization for health behavior change means involving the support from private (e.g., families and friends) or public social networks in health interventions. We have listed the social support types in the review studies in Table 9.5. Katule et al. targeted parent-child pairs to improve the parent’s physical activity via the child (Katule et al., 2016). Chen and colleagues (Y. Chen, Randriambelonoro, Geissbuhler, & Pu, 2016) found that collaborating with a buddy (dyads) to compete in a community can be effective to improve the daily steps of obese and diabetic patients. While studies have shown that social incentives have the potential to
motivate people for health behavior change, some research showed cautions when deploying socialization. E.g., social interactions could be demotivating between dyads who did not know each other well (Y. Chen et al., 2016). The work of Munson et al. (Munson et al., 2015) showed that the prospect of public accountability might suppress the making of commitments which decreases social members’ motivation of obese adults.

More research on the requirements of socialization for different user groups is needed. For instance, how to apply social support to improve physical activity and dietary behavior of the elderly living along?

9.6.6 Embracing AI

Although recent applications of deep learning have boomed in many fields, its use for health behavior change is still in infancy. By AI in health behavior change, we refer to the system that adopts a social role in communicating with users for health behavior change. This definition emphasizes the interaction between the system and users. It excludes the system only providing functional support, e.g., food and ingredient recognition (J. Chen & Ngo, 2016). We found three systems following our definition of AI in health behavior change. Kanaoka and Mutlu used a humanoid robot to motivate users for physical activity in two interaction conditions (Kanaoka & Mutlu, 2015). Over a two-week study, although no significant difference was found in the users’ physical activity level, their intrinsic motivation was significantly improved. Interestingly, users’ willingness and perceived friendliness of the robot are both higher in the monologue condition (less interaction) than in the motivational interviewing condition (more interaction). This result might be due to the lack of fluency and error in speech recognition (Kanaoka & Mutlu, 2015). Another system by Lisetti and colleagues developed a virtual agent to deliver interventions on excessive alcohol consumption (Lisetti, Amini, Yasavur, & Rishe, 2013). Their virtual agent can recognize users’ expressions and generate corresponding expressions to show empathic feedbacks. With empathic feedbacks, the virtual agent improved users’ attitude to the technology, intention to use, perceived enjoyment, and so on. Unlike the mentioned two systems with anthropomorphism, Kamphorst and colleagues developed an autonomous e-coaching system to deliver intervention messages to promote more stairs taking according to the problematic constructs for behavior change of users in real-time (Kamphorst et al., 2014). A month-long evaluation study showed that the intelligent e-coaching system could better support health behavior change.

The initial results of applying AI technology implicate the venues of research on health behavior change in HCI: natural language based intervention (Kanaoka & Mutlu, 2015; Lisetti et al., 2013), emotion enabled intervention (Fadhil, Schiavo, Wang, & Yilma, 2018; Lisetti et al., 2013), and computational intervention (Kamphorst et al., 2014).
9.7 Limitations

Several limitations emerged during the systematic review. The search was conducted only in the ACM digital library, so the reviewed papers did not include all the related work in the HCI community. This might lead to bias in our results. We only searched the authors’ keywords to extract papers, which may also lead to missing some related papers. No paper explicitly reported the intervention strategies based on the taxonomy of behavior change techniques (BCTs). The coding is based on the authors’ understanding of the material provided on the website of BCTs taxonomy\(^\text{25}\), which might introduce bias to our analysis.

9.8 Conclusion

Through a systematic review of the research on health behavior change in the HCI community, this chapter shows the research trends in the perspectives of the target behavior and paper types over the research history (N=648). Based on the selected intervention studies (N=75), it also analyzes the research patterns in the views of measurements, user experience evaluation, the target behavior, the target user group, the use of behavioral theories, the use of behavior change strategies, and intervention characteristics. The results show that physical activity was the most targeted behavior over time, and the related research keeps growing in recent years. Other behaviors (e.g., sleep, dietary behavior, smoking, and sedentary behavior) increasingly draw more attention with slight fluctuations. Studies using interviews or surveys continue to increase, while research on data mining and designing new intervention systems or methods dropped in 2017. Among the 75 intervention studies with compared conditions, only 32\% of the studies quantitatively evaluated the user experience. The SUS, the NASA-TLX, and the AttrakDiff were used in the reviewed studies, while no standardized method to assess the user experience of intervention studies for health behavior change was found. Most of the target users in these studies were adults. There were 32 out of 75 studies explicitly reporting the use of behavioral theory, and the most used one is the transtheoretical model. A variety of behavior change strategies were used in the reviewed studies, while the most frequently used ones include self-monitoring of behavior, goal-setting (behavior), feedback on behavior, prompts/cues, and game integration. Almost all the studies with game integration were designed for physical activity. The mobile phone was the most popular medium to deliver interventions. The plain text, the progress bar, and the gamification interface were the top-3 visualization methods to provide information to the user. Regarding social support, we found nine use cases among the reviewed studies. Based on these findings, we discuss the shortcomings and problems to be addressed: unbalanced target behaviors, unbalanced target user groups, the lack of standardized evaluation methods for

\(^{25}\) http://www.bct-taxonomy.com/dashboard
user experience, and the lack of standards to report intervention studies for health behavior change in HCI.

Finally, we provide suggestions and opportunities for the future research in the field of health behavior change in HCI. We suggest considering users’ behavior priority and the ways to categorize target users when recruiting the study participants. We also point out the lack of longitudinal studies for game-integrated systems and the cautions for socialization-engaged systems. Also, we show how AI technologies have been used for health behavior change and implicate the research venues of AI in this field.

Responding to the trend of decrease in the related paper amount from the HCI community, we believe it is a temporary phenomenon. According to the findings and analysis in this chapter, there are many unexplored research questions and opportunities in health behavior change for HCI researchers.
9.8 Conclusion
Chapter 10: General Discussion

10 General Discussion

"会当凌绝顶，一览众山小。"

We should ascend the mountain’s crest;
It dwarfs all peaks under our feet.¹

- 杜甫 (712 ~ 770)
DU Fu
(A prominent Chinese poet of the Tang dynasty)
¹The translation is adapted from the version by Yuanchong Xu (许渊冲).
10.1 Summary of Contributions

This dissertation contributes to the knowledge body of HCI, health behavior change, and data mining. It reports on the first systematic review to analyze persuasive technologies used in sedentary behavior change intervention studies. We found that reminders were the most frequently used technology. The analysis on reminders showed that hourly PC reminders alone have no significant effect on reducing sedentary behavior at work, while coupling with education or other informative session seems to be promising. Besides, we found several pitfalls in the domain of digital health interventions. Therefore, for the first time to our knowledge, our proposed TUDER framework integrates taxonomies from health psychology, public health, and HCI to enable a consistent and comprehensive approach for designing and reporting digital health interventions. Based on the proposed framework, we present another systematic review of health behavior change research in the HCI community, which provides the latest overview in this domain. Through this review, we analyzed the research trends and patterns regarding the target behaviors/user groups, metrics of evaluation, adopted behavioral theories, and characteristics of interventions.

From a technical perspective, we developed two practical software tools (i.e., SedVis and SedBar) for reducing sedentary behavior. Underlying the tools, we developed data collection and data mining approaches for human mobility patterns analysis. Our clustering algorithm is the first one utilizing the temporal constraints to extract places of interest in spatio-temporal data of human mobility. Our proposed algorithm could extract POIs in the finest spatial granularity with pre-defined temporal constraints (i.e., visiting duration and frequency). We used this clustering algorithm in SedVis and evaluated it using public available large-scale datasets as well as the participants’ data in our study. Our study on the next-place prediction is also the first one to compare the strategy of combining sequential and temporal patterns with those using the patterns separately. Our primary findings include: (1) integrating multiple patterns is not necessarily more effective than using single patterns in average prediction accuracy, and (2) most of the tested methods can outperform others for a specific period (either for the prediction of all places or each place individually). The next-place prediction was evaluated for future work, but not used in our intervention studies reported in this thesis.

Our intervention studies on sedentary behavior change not only evaluated our intervention tools but also provided empirical evidence to examine the behavioral theories. In the first study based on SedVis, we answered three research questions: (1) if visualizations of users’ mobility patterns affect users’ action planning for their sedentary behavior change, (2) if the intervention involving visualizations and action planning is effective in reducing sedentary behavior, (3) if users’ engagement with the visualizations relates to their sedentary behavior change. Our results supported Maher and Conroy’s previous...
finding (Maher & Conroy, 2015): daily action planning alone has no effect on reducing sedentary behavior in the short term among college students. Furthermore, based on the evaluation of participants-made action plans, we found that sedentary behavior change had no correlation with the number of total action plans, the number of unique action plans, the viability of action plans, and the instrumentality of action plans (with weak to moderate evidence).

In the second study, we evaluated two context-aware PC regarding the effect on users’ sedentary behavior change and their perceived usefulness and interruption. One reminder, as the control condition, is the point-of-choice prompt, while the other is our designed SedBar, a self-reflection and reminding tool based on the progress bar. Both reminders used the keyboard/mouse events and the webcam to detect users’ presence as the timing context. So the compared factor was the visualizations of these two reminders. The results showed that more participants preferred SedBar. The participants’ perceived interruption and usefulness also suggested SedBar was more popular during the study. However, the logged data of the participants’ working sessions indicated the prompt was more effective in reducing their sedentary behavior.

The exploration study on the potential impact of augmented-reality head-mounted displays (AR-HMDs) on the movement behavior, to our knowledge, is the first study using the virtual screen to provide different levels of freedom and flexibility of movement. We assumed that the fixed PC screens primarily cause office workers’ sedentary behavior. Then we wanted to know if they will move more if we free the spatial limitation of the physical screens. Therefore, besides the interventions studies leveraging smartphones and PCs, we also explored the potential impact of the augment-reality head-mounted displays (e.g., HoloLens) on office workers’ movement behavior. We compared the differences of the participants’ macro-movements (e.g., sit-stand transitions) and micro-movements (e.g., moving the head) between two experimental modes (i.e., spatial-mapping and tag-along) with different level of movement freedom and flexibility based on a dedicated trivial quiz task we developed in HoloLens. Our lab study revealed some interesting findings: strong evidence supports that participants had more head-movements in the tag-along mode where higher simplicity and freedom of moving the virtual screen were given; body position/direction changes show the same effect with moderate evidence, while sit-stand transitions show no difference between the two modes with weak evidence. Our results also implied several design considerations and research opportunities for future work on the ergonomics of AR-HMDs in the perspective of health.
10.2 Implications and Future Work

10.2.1 The Necessity of More Digital Intervention Studies for Sedentary Behavior Change

Despite the limitations of the systematic reviews in Chapter 2 and 9, they suggested the necessity of more digital intervention studies for sedentary behavior change. From the 1025 initially searched papers, only eight studies met our criteria (see Chapter 2 for details). Compared with other target behaviors (e.g., physical activity and eating behavior), the sedentary behavior was less studies (see Chapter 9 for details). Given the prevalence of sedentary behavior and its contribution to many chronic diseases, well-designed digital intervention studies for reducing prolonged sedentary behavior are still needed to understand the mechanisms of the behavior change. In this dissertation, we only conducted two theory-based empirical studies, which focused on action planning (see Chapter 6 for details) and cues/triggers (see Chapter 7 for details), respectively. Other constructs (e.g., self-efficacy and social pressure, as shown in Figure 3.1) also require experiments to study, which could eventually enable us to answer questions, e.g., how much the existing behavioral theories can explain sedentary behavior change. Therefore, we encourage interdisciplinary cooperation in this field. To this end, importantly, future digital intervention studies should explicitly and formally report the underlying behavior change theories/constructs and intervention strategies including the characteristics (e.g., the media, user interface, and workflow). Our proposed TUDER framework (Chapter 3) could be used for this purpose.

For HCI researchers, another essential question should be answered in future work: how to evaluate the digital health interventions (i.e., not limited to sedentary behavior change). As shown in Figure 9.6, of the 75 reviewed studies, the target behavior was measured in 93% of them, 59% of them evaluated user experience quantitatively or qualitatively, the user interaction with the intervention system was measured in 34% of the studies, and 20% of them accessed users’ behavioral factors/theories. Due to the complexity of human behavior, we suggest future studies to measure and report all the mentions metrics if applicable. For the evaluation of user experience, we found that although about 59% of the studies accessed the user experience, only 32% (24/75) of them evaluated the user experience with formal scales, including the System Usability Scale (SUS), the NASA-TLX and the AttrakDiff. We encourage future studies to evaluate the user experience of their digital interventions and investigate the related factors impacting users’ acceptance of the related technology, as well as users’ behavior change. For example, our study using UEQ (see details in Chapter 6) saw a difference of perceived novelty between two intervention conditions, while a difference in users’ sedentary behavior change was also observed. The relationship between these differences should be further studied.
10.2.2 Understanding and Leveraging Users’ Context

In this dissertation, we illustrated that both the reminders on PCs and visualizations of mobility patterns on smartphones could be served as light-weight digital interventions for reducing the prolonged sedentary behaviors of office workers and college students. In both studies, we used users’ behavioral context to enable more precise and personalized interventions. In SedVis (see Chapter 6 for details), we used users’ mobility patterns for visualizations; In SedBar (see Chapter 7 for details), we used users’ working period to inform the reminders. Through the SedHolo study (see Chapter 8 for details), we also illustrated the necessity of considering users’ movement behavior as the context for designing applications for augmented-reality head-mounted displays.

Besides the users’ context of their behaviors (e.g., mobility patterns, the working period, and postures), other contexts like users’ mood and emotion (Wahl, 2019) could also be essential to the success of interventions. Hence, new technologies, especially the ones that could automatically detect users’ context during the deployed intervention, are required to enable better design of digital health interventions. For example, the concept of stress detection at work proposed by Ferreira et al. (Ferreira et al., 2014) should be realized in future work.

10.2.3 Enhancing Users’ Interaction and Engagement

The results of the intervention study based on SedVis (see Chapter 6 for details) implied the importance of users’ engagement in digital health interventions. As shown in Figure 6.10, users’ engagement was positively related to the effect of reducing sedentary hours. However, it is unknown which factors contribute to users’ engagement, which possibly include the perceived usefulness, aesthetics of the app, social network, to name a few. Therefore, future work might investigate the mechanism underlying users’ engagement with digital health interventions.

Specifically, we encourage future work to study the effect of users’ participation in the design/assembly of intervention tools on users’ engagement. In our prior work based on the concept of the IKEA effect (Norton et al., 2012; Y. Wang, Pfeil, et al., 2016), we hypothesized that users’ participation in the design/assembly process of making the digital intervention tools could moderate users’ engagement. Given that the hypothesis is true, how can we design the users’ participation process to maximize the effect? These questions should be answered in future work.

Another research direction we would like to promote is studying the necessity/effect of the intelligibility of digital health intervention systems on users’ acceptance and engagement. Intelligibility is used as a property of context-aware systems indicating the ability to represent to their users what they know, how they know it, and what they are doing about it (Bellotti & Edwards, 2001). We follow this definition in the context of digital health interventions. Our
10.2 Implications and Future Work

A proposal to study the intelligibility is due to its possible correlation with users’ acceptance and engagement with digital health intervention tools. The ultimate goal of health behavior change systems is to improve users’ unhealthy behavior/habits effectively, efficiently, and nonobtrusively. In reality, most such systems have been abandoned before they affect (T. Fritz, Huang, Murphy, & Zimmermann, 2014; Gouveia, Karapanos, & Hassenzahl, 2015; Meyer, Wasmann, Heuten, El Ali, & Boll, 2017). Taking physical activity for example (Meyer et al., 2017), many reasons for user abandonment have been discussed: cost of collecting, cost of ownership, discomfort with information, data quality concerns, learned enough, change in life circumstances, expectation mismatch, technical complexity, etc. The lack of intelligibility (e.g., expectation mismatch) could be an unnoticeable reason. Recalling the result of our systematic review of sedentary behavior change in Chapter 2: interventions combining education and PC reminders seem to work better than only PC reminders for reducing sedentary behavior. The education here served as a role of explanation. Therefore, we think intelligibility is an important but unexplored topic for health behavior change in HCI. Hence, the impact of intelligibility of health behavior change systems on users’ engagement, long-term acceptance, and behavior change should be further investigated.

10.2.4 Regarding Users’ Health Behavior Change in A Holistic View

Human behaviors and activities are complex and correlated. A person’s daily life/routine consists of a list of interdependent behaviors/activities. For example, if a person does not sleep well during the night, her/his eating and physical activity might be affected subsequently. We notice that most health intervention studies, including ours, only targeted a single behavior and neglected others. The aim, consequently, was to change the target behavior, temporarily at lease. We argue that single-behavior thinking might be problematic from the perspective of holistic health, because the neglected behaviors related to health might simultaneously change.

Inspired by the philosophy of holistic medicine (WebMD, 2017), we propose to regard users’ health behavior change in a holistic view, which refers to considering all the major behaviors related to a user when targeting one behavior. The major behaviors include the leading health-risk behaviors (WHO, 2018c) (i.e., smoking, drinking alcohol, physical inactivity, unbalanced diet) and other evidence-based harmful behaviors (e.g., prolonged sedentary behavior, insufficient water intake (Popkin, D’Ancí, & Rosenberg, 2010), and insufficient sleep (Hershner & Chervin, 2014)). Digital health interventions adopting the holistic view should assess, analyze, and report the major behaviors before and after the intervention, in addition to the target behavior.
REFERENCES


References


morbidities or additional risk factors for co-morbidities: a systematic review. *Health Psychology Review*, 6(1), 7–32. https://doi.org/10.1080/17437199.2010.513298


References


References

Incremental Parsing: The Active LeZi Algorithm. *IEEE Intelligent Systems, 22*(1), 52–58. https://doi.org/10.1109/MIS.2007.15


References


References

https://doi.org/10.1016/j.jbi.2015.06.001


Khoroshevsky, F., & Lerner, B. (2016). Human Mobility-Pattern Discovery and Next-Place Prediction from GPS Data. In MPRSS (pp. 24–35). https://doi.org/10.1007/978-3-319-59259-6

References


References


on Interaction Design and Children (pp. 60–69). New York, NY, USA: ACM. https://doi.org/10.1145/2307096.2307104


Interventions to Change Health Behavior. *Journal of Medical Internet Research, 18*(12), e317. https://doi.org/10.2196/jmir.5927


References


References

https://affect.media.mit.edu/pdfs/17.sano_etal_stressinterventiondelivery_ACII.pdf


References


References

https://doi.org/10.1016/j.smhl.2018.05.002


References


This table presents the PSD principles and BCTs, where the column "PSD -> BCT" indicates the overlapped items.

<table>
<thead>
<tr>
<th>PSD principles</th>
<th>PSD principle explanation</th>
<th>PSD -&gt; BCT</th>
<th>BCTs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Task Support</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reducione (1.1)</td>
<td>System should reduce steps users take when performing target behavior.</td>
<td>1.1 -&gt; 1.2</td>
<td>1.1. Goal setting (behavior)</td>
</tr>
<tr>
<td>Tunneling (1.2)</td>
<td>System should guide users in attitude/ behavior change process by providing means for action.</td>
<td>1.2 -&gt; 4.1</td>
<td>1.2. Problem solving</td>
</tr>
<tr>
<td>Tailoring (1.3)</td>
<td>System should provide tailored info for user groups.</td>
<td></td>
<td>1.3. Goal setting (outcome)</td>
</tr>
<tr>
<td>Personality (1.4)</td>
<td>System should offer personalized content and services for individual users.</td>
<td></td>
<td>1.4. Action planning</td>
</tr>
<tr>
<td>Self-monitoring (1.5)</td>
<td>System should provide means for users to track their performance or status.</td>
<td>1.5 -&gt; 2.3</td>
<td>1.5. Review behavior goal(s)</td>
</tr>
<tr>
<td>Simulations (1.6)</td>
<td>System should provide means for observing link between cause &amp; effect with regard to users’ behavior.</td>
<td>1.6 -&gt; 4.3</td>
<td>1.6. Discrepancy between current behavior and goal</td>
</tr>
<tr>
<td>Rehearsal (1.7)</td>
<td>System should provide means for rehearsing target behavior.</td>
<td>1.7 -&gt; 8.1</td>
<td>1.7. Review outcome goal(s)</td>
</tr>
<tr>
<td><strong>Dialogue Support</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prais (2.1)</td>
<td>System should use praise to provide user feedback based on behaviors.</td>
<td>2.1 -&gt; 10.4</td>
<td>1.8. Behavioral contract</td>
</tr>
<tr>
<td>Rewards (2.2)</td>
<td>System should provide virtual rewards for users to give credit for performing target behavior.</td>
<td>2.2 -&gt; 10.3</td>
<td>1.9. Commitment</td>
</tr>
<tr>
<td>Reminders (2.3)</td>
<td>System should remind users of their target behavior while using the system.</td>
<td>2.3 -&gt; 7.1</td>
<td>2.1. Monitoring of behavior by others without feedback</td>
</tr>
<tr>
<td>Suggestion (2.4)</td>
<td>System should suggest users carry out behaviors while using the system.</td>
<td>2.4 -&gt; 4.1</td>
<td>2.2. Feedback on behaviour</td>
</tr>
<tr>
<td>Similarity (2.5)</td>
<td>System should imitate its users in some specific way.</td>
<td></td>
<td>2.3. Self-monitoring of behaviour</td>
</tr>
<tr>
<td>Liking (2.6)</td>
<td>System should have a look &amp; feel that appeals to users.</td>
<td></td>
<td>2.4. Self-monitoring of outcome(s) of behaviour</td>
</tr>
<tr>
<td>Social role (2.7)</td>
<td>System should adopt a social role.</td>
<td></td>
<td>2.5. Monitoring of outcome(s) of behavior without feedback</td>
</tr>
<tr>
<td><strong>System Credibility Support</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trustworthiness (3.1)</td>
<td>System should provide info that is truthful, fair &amp; unbiased.</td>
<td></td>
<td>2.6. Biofeedback</td>
</tr>
<tr>
<td>Expertise (3.2)</td>
<td>System should provide info showing knowledge, experience &amp; competence.</td>
<td>3.2 -&gt; 5.6</td>
<td>2.7. Feedback on outcome(s) of behavior</td>
</tr>
<tr>
<td>Surface credibility (3.3)</td>
<td>System should have competent and truthful look &amp; feel.</td>
<td></td>
<td>3.1. Social support (unspecified)</td>
</tr>
<tr>
<td>Real-world feel (3.4)</td>
<td>System should provide info of the organization/actual people behind it content &amp; services.</td>
<td>3.4 -&gt; 9.1</td>
<td>3.2. Social support (practical)</td>
</tr>
<tr>
<td>Authority (3.5)</td>
<td>System should refer to people in the role of authority.</td>
<td>3.5 -&gt; 9.1</td>
<td>3.3. Social support (emotional)</td>
</tr>
<tr>
<td>Third-party endorsements (3.6)</td>
<td>System should provide endorsements from external sources.</td>
<td>3.6 -&gt; 9.1</td>
<td>4. Shaping knowledge</td>
</tr>
<tr>
<td>Verifiability (3.7)</td>
<td>System should provide means to verify accuracy of site content via outside sources.</td>
<td>3.7 -&gt; 9.1</td>
<td>4.1. Instruction on how to perform the behavior</td>
</tr>
</tbody>
</table>

**APPENDIX 1**
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Appendix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social learning (4.1)</td>
<td>System should provide means to observe others performing their target behaviors.</td>
<td>4.1 -&gt; 3.2</td>
</tr>
<tr>
<td>Social comparison (4.2)</td>
<td>System should provide means for comparing performance with the performance of others.</td>
<td>4.2 -&gt; 6.2</td>
</tr>
<tr>
<td>Normative influence (4.3)</td>
<td>System should provide means for gathering people who have same goal &amp; make them feel norms.</td>
<td>4.3 -&gt; 6.3</td>
</tr>
<tr>
<td>Social facilitation (4.4)</td>
<td>System should provide means for discerning others who are performing the behavior.</td>
<td>4.4 -&gt; 6.3</td>
</tr>
<tr>
<td>Cooperation (4.5)</td>
<td>System should provide means for cooperation.</td>
<td></td>
</tr>
<tr>
<td>Competition (4.6)</td>
<td>System should provide means for competing with others.</td>
<td></td>
</tr>
<tr>
<td>Recognition (4.7)</td>
<td>System should provide public recognition for users who perform their target behavior.</td>
<td></td>
</tr>
<tr>
<td>5. Natural consequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1. Information about health consequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2. Salience of consequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.3. Information about social and environmental consequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.4. Monitoring of emotional consequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5. Anticipated regret</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.6. Information about emotional consequences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Comparison of behaviour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1. Demonstration of the behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2. Social comparison</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3. Information about other’s approval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Associations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.1. Prompts/cues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.2. Cue signalling reward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.3. Reduce prompts/cues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.4. Remove access to the reward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.5. Remove aversive stimulus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.6. Satiation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.7. Exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.8. Associative learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Repetition and substitution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.1. Behavioral practice/rehearsal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.2. Behavior substitution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.3. Habit formation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.4. Habit reversal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.5. Overcorrection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.6. Generalisation of target behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.7. Graded tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Comparison of outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.1. Credible source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.2. Pros and cons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.3. Comparative imagining of future outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Reward and threat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.1. Material incentive (behavior)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.2. Material reward (behavior)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.3. Non-specific reward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.4. Social reward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.5. Social incentive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.6. Non-specific incentive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.7. Self-incentive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Appendix 1</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>10.8.</td>
<td>Incentive (outcome)</td>
<td></td>
</tr>
<tr>
<td>10.9.</td>
<td>Self-reward</td>
<td></td>
</tr>
<tr>
<td>10.10.</td>
<td>Reward (outcome)</td>
<td></td>
</tr>
<tr>
<td>10.11.</td>
<td>Future punishment</td>
<td></td>
</tr>
<tr>
<td><strong>11. Regulation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.1.</td>
<td>Pharmacological support</td>
<td></td>
</tr>
<tr>
<td>11.2.</td>
<td>Reduce negative emotions</td>
<td></td>
</tr>
<tr>
<td>11.3.</td>
<td>Conserving mental resources</td>
<td></td>
</tr>
<tr>
<td>11.4.</td>
<td>Paradoxical instructions</td>
<td></td>
</tr>
<tr>
<td><strong>12. Antecedents</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.1.</td>
<td>Restructuring the physical environment</td>
<td></td>
</tr>
<tr>
<td>12.2.</td>
<td>Restructuring the social environment</td>
<td></td>
</tr>
<tr>
<td>12.3.</td>
<td>Avoidance/reducing exposure to cues for the behavior</td>
<td></td>
</tr>
<tr>
<td>12.4.</td>
<td>Distraction</td>
<td></td>
</tr>
<tr>
<td>12.5.</td>
<td>Adding objects to the environment</td>
<td></td>
</tr>
<tr>
<td>12.6.</td>
<td>Body changes</td>
<td></td>
</tr>
<tr>
<td><strong>13. Identity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.1.</td>
<td>Identification of self as role model</td>
<td></td>
</tr>
<tr>
<td>13.2.</td>
<td>Framing/reframing</td>
<td></td>
</tr>
<tr>
<td>13.3.</td>
<td>Incompatible beliefs</td>
<td></td>
</tr>
<tr>
<td><strong>14. Scheduled consequences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.1.</td>
<td>Behavior cost</td>
<td></td>
</tr>
<tr>
<td>14.2.</td>
<td>Punishment</td>
<td></td>
</tr>
<tr>
<td>14.3.</td>
<td>Remove reward</td>
<td></td>
</tr>
<tr>
<td>14.4.</td>
<td>Reward approximation</td>
<td></td>
</tr>
<tr>
<td>14.5.</td>
<td>Rewarding completion</td>
<td></td>
</tr>
<tr>
<td>14.6.</td>
<td>Situation-specific reward</td>
<td></td>
</tr>
<tr>
<td>14.7.</td>
<td>Reward incompatible behavior</td>
<td></td>
</tr>
<tr>
<td>14.8.</td>
<td>Reward alternative behavior</td>
<td></td>
</tr>
<tr>
<td>14.9.</td>
<td>Reduce reward frequency</td>
<td></td>
</tr>
<tr>
<td>14.10.</td>
<td>Remove punishment</td>
<td></td>
</tr>
<tr>
<td><strong>15. Self-belief</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.1.</td>
<td>Verbal persuasion about capability</td>
<td></td>
</tr>
<tr>
<td>15.2.</td>
<td>Mental rehearsal of successful performance</td>
<td></td>
</tr>
<tr>
<td>15.3.</td>
<td>Focus on past success</td>
<td></td>
</tr>
<tr>
<td>15.4.</td>
<td>Self-talk</td>
<td></td>
</tr>
<tr>
<td><strong>16. Covert learning</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.1.</td>
<td>Imaginary punishment</td>
<td></td>
</tr>
<tr>
<td>16.2.</td>
<td>Imaginary reward</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX 2

The tables present the taxonomy of digital health interventions (DHIs), which includes the strategy part and the characteristic part. The strategy part includes 96 items, of which 93 are from the BCT taxonomy while 3 are from the PSD principles (3.4-3.6 in green). The characteristic part includes 6 items, of which 4 are from the BIT model while 2 are extracted from the PSD principles (in blue). The strategy description of the items from BCT taxonomy is adapted from the material in the BCT-taxonomy website (http://www.bct-taxonomy.com/). The characteristics adapted from the BIT model might be differently explained in our DHI taxonomy.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Characteristic Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium</td>
<td>The media through which an intervention is delivered. It includes both the hardware (e.g., PC, smartphones, wearables) and the user interface (e.g., text, video, audio).</td>
</tr>
<tr>
<td>Complexity</td>
<td>To what extent the intervention media is complex for users to interact.</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>To what extent the intervention media presents users aesthetics.</td>
</tr>
<tr>
<td>Personalization</td>
<td>To what extent the intervention adapts to users’ preferences.</td>
</tr>
<tr>
<td>Social role</td>
<td>The role the intervention plays when providing social support to users (e.g., social sharing, social comparison, social competition, social communication, social monitoring).</td>
</tr>
<tr>
<td>Trustiness</td>
<td>To what extent the intervention provides means to increase users’ trustiness.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Strategy Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Goals and planning</td>
<td></td>
</tr>
<tr>
<td>1.1. Goal setting (behavior)</td>
<td>Set or agree on a goal defined in terms of the behavior to be achieved.</td>
</tr>
<tr>
<td>1.2. Problem solving</td>
<td>Analyse, or prompt the person to analyse, factors influencing the behavior and generate or select strategies that include overcoming barriers and/or increasing facilitators (includes 'Relapse Prevention' and 'Coping Planning'). Note: barrier identification without solutions is not sufficient. If the BCT does not include analysing the behavioral problem, consider 12.3, Avoidance/changing exposure to cues for the behavior, 12.1, Restructuring the physical environment, 12.2, Restructuring the social environment, or 11.2, Reduce negative emotions.</td>
</tr>
<tr>
<td>1.3. Goal setting (outcome)</td>
<td>Set or agree on a goal defined in terms of a positive outcome of wanted behavior.</td>
</tr>
<tr>
<td>1.4. Action planning</td>
<td>Prompt detailed planning of performance of the behavior (must include at least one of context, frequency, duration and intensity). Context may be environmental (physical or social) or internal (physical, emotional or cognitive)(includes 'Implementation Intentions'). Note: evidence of action planning does not necessarily imply goal setting, only code latter if sufficient evidence.</td>
</tr>
<tr>
<td>1.5. Review behavior goal(s)</td>
<td>Review behavior goal(s) jointly with the person and consider modifying goal(s) or behavior change strategy in light of achievement. This may lead to re-setting the same goal, a small change in that goal or setting a new goal instead of (or in addition to) the first, or no change. Note: if goal specified in terms of behavior, code 1.5, Review behavior goal(s), if goal unspecified, code 1.7, Review outcome goal(s); if discrepancy created consider also 1.6, Discrepancy between current behavior and goal.</td>
</tr>
<tr>
<td>1.6. Discrepancy between current behavior and goal</td>
<td>Draw attention to discrepancies between a person’s current behavior (in terms of the form, frequency, duration, or intensity of that behavior) and the person’s previously set outcome goals, behavioral goals or action plans (goes beyond self-monitoring of behavior) Note: if discomfort is created only code 13.3, Incompatible beliefs and not 1.6, Discrepancy between current behavior and goal; if goals are modified, also code 1.5, Review behavior goal(s) and/or 1.7, Review outcome goal(s); if feedback is provided, also code 2.2, Feedback on behaviour.</td>
</tr>
</tbody>
</table>
1.7. Review outcome goal(s) | Review outcome goal(s) jointly with the person and consider modifying goal(s) in light of achievement. This may lead to re-setting the same goal, a small change in that goal or setting a new goal instead of, or in addition to the first. Note: if goal specified in terms of behavior, code 1.5, Review behavior goal(s); if goal unspecified, code 1.7, Review outcome goal(s); if discrepancy created consider also 1.6, Discrepancy between current behavior and goal.

1.8. Behavioral contract | Create a written specification of the behavior to be performed, agreed on by the person, and witnessed by another. Note: also code 1.1, Goal setting (behavior).

1.9. Commitment | Note: if defined in terms of the behavior to be achieved also code 1.1, Goal setting (behavior)

2. Feedback and monitoring

2.1. Monitoring of behavior by others without feedback | Observe or record behavior with the person’s knowledge as part of a behavior change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behavior, do not code; if feedback given, code only 2.2, Feedback on behavior, and not 2.1. Monitoring of behavior by others with feedback; if monitoring outcome(s) code 2.5, Monitoring outcome(s) of behavior by others without feedback; if self-monitoring behavior, code 2.3, Self-monitoring of behavior.

2.2. Feedback on behaviour | Monitor and provide informative or evaluative feedback on performance of the behavior (e.g. form, frequency, duration, intensity). Note: if Biofeedback, code only 2.6, Biofeedback and not 2.2, Feedback on behavior; if feedback is on outcome(s) of behavior, code 2.7, Feedback on outcome(s) of behavior; if there is no clear evidence that feedback was given, code 2.1, Monitoring of behavior by others without feedback; if feedback on behaviour is evaluative e.g. praise, also code 10.4, Social reward.

2.3. Self-monitoring of behaviour | Establish a method for the person to monitor and record their behavior(s) as part of a behavior change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behavior, do not code; if monitoring of outcome of behavior, code 2.4, Self-monitoring of outcome(s) of behavior; if monitoring is by someone else (without feedback), code 2.1, Monitoring of behavior by others without feedback.

2.4. Self-monitoring of outcome(s) of behaviour | Establish a method for the person to monitor and record the outcome(s) of their behavior as part of a behavior change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behavior, do not code; if monitoring behavior, code 2.3, Self-monitoring of behavior; if monitoring is by someone else (without feedback), code 2.5, Monitoring outcome(s) of behavior by others without feedback.

2.5. Monitoring of outcome(s) of behavior without feedback | Observe or record outcomes of behavior with the person’s knowledge as part of a behavior change strategy. Note: if monitoring is part of a data collection procedure rather than a strategy aimed at changing behavior, do not code; if feedback given, code only 2.7, Feedback on outcome(s) of behavior; if monitoring behavior code 2.1, Monitoring of behavior by others without feedback; if self-monitoring outcome(s), code 2.4, Self-monitoring of outcome(s) of behavior.

2.6. Biofeedback | Provide feedback about the body (e.g. physiological or biochemical state) using an external monitoring device as part of a behavior change strategy. Note: if Biofeedback, code only 2.6, Biofeedback and not 2.2, Feedback on behavior or 2.7, Feedback on outcome(s) of behavior.

2.7. Feedback on outcome(s) of behavior | Monitor and provide feedback on the outcome of performance of the behavior. Note: if Biofeedback, code only 2.6, Biofeedback and not 2.7, Feedback on outcome(s) of behavior; if feedback is on behavior code 2.2, Feedback on behavior; if there is no clear evidence that feedback was given code 2.5, Monitoring outcome(s) of behavior by others without feedback; if feedback on behaviour is evaluative e.g. praise, also code 10.4, Social reward.

3. Social support

3.1. Social support (unspecified) | Advise on, arrange or provide social support (e.g. from friends, relatives, colleagues,' buddies' or staff) or non-contingent praise or reward for performance of the behavior. It includes encouragement and counselling, but only when it is directed at the behavior.
### Appendix 2

#### 3.2. Social support (practical)
Advising on, arranging, or providing practical help (e.g. from friends, relatives, colleagues, ‘buddies’ or staff) for performance of the behavior.  
**Note:** if emotional, code 3.3, Social support (emotional); if general or unspecified, code 3.1, Social support (unspecified). If only restructuring the physical environment or adding objects to the environment, code 12.1, Restructuring the physical environment or 12.5, Adding objects to the environment; attending a group or class and/or mention of ‘follow-up’ does not necessarily apply this BCT, support must be explicitly mentioned.

#### 3.3. Social support (emotional)
Advising on, arranging, or providing emotional social support (e.g. from friends, relatives, colleagues, ‘buddies’ or staff) for performance of the behavior.  
**Note:** if practical, code 3.2, Social support (practical); if unspecified, code 3.1, Social support (unspecified).

#### 3.4 Cooperation
Provide means for cooperation with others.

#### 3.5 Competition
Provide means for competing with others.

#### 3.6 Recognition
Provide public recognition for users who perform their target behavior.

### 4. Shaping knowledge

#### 4.1. Instruction on how to perform the behavior
Advising or agreeing on how to perform the behavior (includes ‘Skills training’).  
**Note:** when the person attends classes such as exercise or cookery, code 4.1, Instruction on how to perform the behavior, 8.1, Behavioral practice/rehearsal and 6.1, Demonstration of the behavior.

#### 4.2. Information about Antecedents
Provide information about antecedents. E.g., social and environmental situations and events, emotions, cognitions that reliably predict performance of the behavior.

#### 4.3. Re-attribution
Elicit perceived causes of behavior and suggest alternative explanations (e.g. external or internal and stable or unstable).

#### 4.4. Behavioral experiments
Advising on how to identify and test hypotheses about the behavior, its causes and consequences, by collecting and interpreting data.

### 5. Natural consequences

#### 5.1. Information about health consequences
Provide information (e.g. written, verbal, visual) about health consequences of performing the behavior.  
**Note:** consequences can be for any target, not just the recipient(s) of the intervention; emphasizing importance of consequences is not sufficient; if information about emotional consequences, code 5.6, Information about emotional consequences; if about social, environmental or unspecified consequences code 5.3, Information about social and environmental consequences.

#### 5.2. Salience of consequences
Use methods specifically designed to emphasize the consequences of performing the behaviour with the aim of making them more memorable (goes beyond informing about consequences).  
**Note:** if information about consequences, also code 5.1, Information about health consequences, 5.6, Information about emotional consequences or 5.3, Information about social and environmental consequences.

#### 5.3. Information about social and environmental consequences
Provide information (e.g. written, verbal, visual) about social and environmental consequences of performing the behavior.  
**Note:** consequences can be for any target, not just the recipient(s) of the intervention; if information about health or consequences, code 5.1, Information about health consequences; if about emotional consequences, code 5.6, Information about emotional consequences; if unspecified, code 5.3, Information about social and environmental consequences.

#### 5.4. Monitoring of emotional consequences
Prompt assessment of feelings after attempts at performing the behavior.

#### 5.5. Anticipated regret
Induce or raise awareness of expectations of future regret about performance of the unwanted behavior.  
**Note:** not including 5.6, Information about emotional consequences if suggests adoption of a perspective or new perspective in order to change cognitions also code 13.2, Framing/reframing.
5.6. Information about emotional consequences

Provide information (e.g. written, verbal, visual) about emotional consequences of performing the behavior.

Note: consequences can be related to emotional health disorders (e.g. depression, anxiety) and/or states of mind (e.g. low mood, stress); not including 5.5, Anticipated regret; consequences can be for any target, not just the recipient(s) of the intervention; if information about health consequences code 5.1, Information about health consequences; if about social, environmental or unspecified code 5.3, Information about social and environmental consequences.

6. Comparison of behaviour

6.1. Demonstration of the behavior

Provide an observable sample of the performance of the behaviour, directly in person or indirectly e.g. via film, pictures, for the person to aspire to or imitate (includes 'Modelling'). Note: if advised to practice, also code, 8.1, Behavioral practice and rehearsal; If provided with instructions on how to perform, also code 4.1, Instruction on how to perform the behaviour.

6.2. Social comparison

Draw attention to others' performance to allow comparison with the person's own performance Note: being in a group setting does not necessarily mean that social comparison is actually taking place.

6.3. Information about other's approval

Provide information about what other people think about the behavior. The information clarifies whether others will like, approve or disapprove of what the person is doing or will do.

7. Associations

7.1. Prompts/cues

Introduce or define environmental or social stimulus with the purpose of prompting or cueing the behavior. The prompt or cue would normally occur at the time or place of performance.

Note: when a stimulus is linked to a specific action in an if-then plan including one or more of frequency, duration or intensity also code 1.4, Action planning.

7.2. Cue signaling reward

Identify an environmental stimulus that reliably predicts that reward will follow the behavior (includes 'Discriminative cue').

7.3. Reduce prompts/cues

Withdraw gradually prompts to perform the behavior (includes 'Fading').

7.4. Remove access to the reward

Advise or arrange for the person to be separated from situations in which unwanted behavior can be rewarded in order to reduce the behavior (includes 'Time out').

7.5. Remove aversive stimulus

Advise or arrange for the removal of an aversive stimulus to facilitate behavior change (includes 'Escape learning').

7.6. Satiation

Advise or arrange repeated exposure to a stimulus that reduces or extinguishes a drive for the unwanted behavior.

7.7. Exposure

Provide systematic confrontation with a feared stimulus to reduce the response to a later encounter.

7.8. Associative learning

Present a neutral stimulus jointly with a stimulus that already elicits the behavior repeatedly until the neutral stimulus elicits that behavior (includes 'Classical/Pavlovian Conditioning').

Note: when a BCT involves reward or punishment, code one or more of: 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).

8. Repetition and substitution

8.1. Behavioral practice/rehearsal

Prompt practice or rehearsal of the performance of the behaviour one or more times in a context or at a time when the performance may or may not be necessary, in order to increase habit and skill.

Note: if aiming to associate performance with the context, also code 8.3, Habit formation.

8.2. Behavior substitution

Prompt substitution of the unwanted behavior with a wanted or neutral behavior.

Note: if this occurs regularly, also code 8.4, Habit reversal.

8.3. Habit formation

Prompt rehearsal and repetition of the behavior in the same context repeatedly so that the context elicits the behavior.

8.4. Habit reversal

Prompt rehearsal and repetition of an alternative behavior to replace an unwanted habitual behavior.

Note: also code 8.2, Behavior substitution.

8.5. Overcorrection

Ask to repeat the wanted behavior in an exaggerated way following an unwanted behaviour.
8.6. Generalization of target behavior
Advising to perform the wanted behavior, which is already performed in a particular situation, in another situation.

8.7. Graded tasks
Set easy-to-perform tasks, making them increasingly difficult, but achievable, until behavior is performed.

9. Comparison of outcomes

9.1. Credible source
Present verbal or visual communication from a credible source in favour of or against the behavior. Note: code this BCT if source generally agreed on as credible e.g., health professionals, celebrities or words used to indicate expertise or leader in field and if the communication has the aim of persuading; if information about health consequences, also code 5.1, Information about health consequences, if about emotional consequences, also code 5.6, Information about emotional consequences; if about social, environmental or unspecified consequences also code 5.3, Information about social and environmental consequences.

9.2. Pros and cons
Advise the person to identify and compare reasons for wanting (pros) and not wanting to (cons) change the behavior (includes 'Decisional balance'). Note: if providing information about health consequences, also code 5.1, Information about health consequences; if providing information about emotional consequences, also code 5.6, Information about emotional consequences; if providing information about social, environmental or unspecified consequences also code 5.3, Information about social and environmental consequences.

9.3. Comparative imagining of future outcomes
Prompt or advise the imagining and comparing of future outcomes of changed versus unchanged behaviour.

10. Reward and threat

10.1. Material incentive (behavior)
Arrange for the delivery of a reward if and only if there has been effort and/or progress in achieving the behavioral outcome (includes 'Positive reinforcement'). Note: this includes social, material, self- and non-specific rewards for outcome; if reward is for the behavior code 10.4, Social reward, 10.2, Material reward (behavior), 10.3, Non-specific reward or 10.9, Self-reward and not 10.10, Reward (outcome). If informed of reward in advance of rewarded behaviour, also code one of: 10.1, Material incentive (behaviour); 10.5, Social incentive; 10.6, Non-specific incentive; 10.7, Self-incentive; 10.8, Incentive (outcome).

10.2. Material reward (behavior)
Arrange for the delivery of money, vouchers or other valued objects if and only if there has been effort and/or progress in performing the behavior (includes 'Positive reinforcement'). Note: if reward is social, code 10.4, Social reward, if unspecified code 10.3, Non-specific reward, and not 10.1, Material reward (behavior); if reward is for outcome, code 10.10, Reward (outcome). If informed of reward in advance of rewarded behaviour, also code one of: 10.1, Material incentive (behaviour); 10.5, Social incentive; 10.6, Non-specific incentive; 10.7, Self-incentive; 10.8, Incentive (outcome).

10.3. Non-specific reward
Arrange delivery of a reward if and only if there has been effort and/or progress in performing the behavior (includes 'Positive reinforcement'). Note: if reward is material, code 10.2, Material reward (behavior), if social, code 10.4, Social reward, and not 10.3, Non-specific reward; if reward is for outcome code 10.10, Reward (outcome). If informed of reward in advance of rewarded behaviour, also code one of: 10.1, Material incentive (behaviour); 10.5, Social incentive; 10.6, Non-specific incentive; 10.7, Self-incentive; 10.8, Incentive (outcome).

10.4. Social reward
Arrange verbal or non-verbal reward if and only if there has been effort and/or progress in performing the behavior (includes 'Positive reinforcement'). Note: if reward is material, code 10.2, Material reward (behavior), if unspecified code 10.3, Non-specific reward, and not 10.4, Social reward; if reward is for outcome code 10.10, Reward (outcome). If informed of reward in advance of rewarded behaviour, also code one of: 10.1, Material incentive (behaviour); 10.5, Social incentive; 10.6, Non-specific incentive; 10.7, Self-incentive; 10.8, Incentive (outcome).
10.5. Social incentive
Inform that a verbal or non-verbal reward will be delivered if and only if there has been effort and/or progress in performing the behavior (includes 'Positive reinforcement').
Note: if incentive is material, code 10.1, Material incentive (behavior), if unspecified code 10.6, Non-specific incentive, and not 10.5, Social incentive; if incentive is for outcome code 10.8, Incentive (outcome). If reward is delivered also code one of: 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).

10.6. Non-specific incentive
Inform that a reward will be delivered if and only if there has been effort and/or progress in performing the behavior (includes 'Positive reinforcement').
Note: if incentive is material, code 10.1, Material incentive (behavior), if social, code 10.5, Social incentive and not 10.6, Non-specific incentive; if incentive is for outcome code 10.8, Incentive (outcome). If reward is delivered also code one of: 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).

10.7. Self-incentive
Plan to reward self in future if and only if there has been effort and/or progress in performing the behavior.
Note: if self-reward is material, also code 10.1, Material incentive (behavior), if social, also code 10.5, Social incentive, if unspecified, also code 10.6, Non-specific incentive; if incentive is for outcome code 10.8, Incentive (outcome). If reward is delivered also code one of: 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).

10.8. Incentive (outcome)
Inform that a reward will be delivered if and only if there has been effort and/or progress in achieving the behavioural outcome (includes 'Positive reinforcement').
Note: this includes social, material, self- and non-specific incentives for outcome; if incentive is for the behavior code 10.5, Social incentive, 10.1, Material incentive (behavior), 10.6, Non-specific incentive or 10.7, Self-incentive and not 10.8, Incentive (outcome). If reward is delivered also code one of: 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).

10.9. Self-reward
Prompt self-praise or self-reward if and only if there has been effort and/or progress in performing the behavior.
Note: if self-reward is material, also code 10.2, Material reward (behavior), if social, also code 10.4, Social reward, if unspecified, also code 10.3, Non-specific reward; if reward is for outcome code 10.10, Reward (outcome). If informed of reward in advance of rewarded behaviour, also code one of: 10.1, Material incentive (behaviour); 10.5, Social incentive; 10.6, Non-specific incentive; 10.7, Self-incentive; 10.8, Incentive (outcome).

10.10. Reward (outcome)
Arrange for the delivery of a reward if and only if there has been effort and/or progress in achieving the behavioral outcome (includes 'Positive reinforcement').
Note: this includes social, material, self- and non-specific rewards for outcome; if reward is for the behavior code 10.4, Social reward, 10.2, Material reward (behavior), 10.3, Non-specific reward or 10.9, Self-reward and not 10.10, Reward (outcome). If informed of reward in advance of rewarded behaviour, also code one of: 10.1, Material incentive (behaviour); 10.5, Social incentive; 10.6, Non-specific incentive; 10.7, Self-incentive; 10.8, Incentive (outcome).

10.11. Future punishment
Inform that future punishment or removal of reward will be a consequence of performance of an unwanted behavior (may include fear arousal) (includes 'Threat').

11. Regulation
11.1. Pharmacological support
Provide, or encourage the use of or adherence to, drugs to facilitate behavior change.
Note: if pharmacological support to reduce negative emotions (i.e. anxiety) then also code 11.2, Reduce negative emotions.

11.2. Reduce negative emotions
Advise on ways of reducing negative emotions to facilitate performance of the behavior (includes 'Stress Management').
Note: if includes analysing the behavioural problem, also code 1.2, Problem solving.

11.3. Conserving mental resources
Advise on ways of minimising demands on mental resources to facilitate behavior change.

11.4. Paradoxical instructions
Advise to engage in some form of the unwanted behavior with the aim of reducing motivation to engage in that behaviour.

200
### 12. Antecedents

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1. Restructuring the physical environment</td>
<td>Change, or advise to change the physical environment in order to facilitate performance of the wanted behavior or create barriers to the unwanted behavior (other than prompts/cues, rewards and punishments). Note: this may also involve 12.3, Avoidance/reducing exposure to cues for the behavior; if restructuring of the social environment code 12.2, Restructuring the social environment; if only adding objects to the environment, code 12.5, Adding objects to the environment.</td>
</tr>
<tr>
<td>12.2. Restructuring the social environment</td>
<td>Change, or advise to change the social environment in order to facilitate performance of the wanted behavior or create barriers to the unwanted behavior (other than prompts/cues, rewards and punishments). Note: this may also involve 12.3, Avoidance/reducing exposure to cues for the behavior; if also restructuring of the physical environment also code 12.1, Restructuring the physical environment.</td>
</tr>
<tr>
<td>12.3. Avoidance/reducing exposure to cues for the behavior</td>
<td>Advise on how to avoid exposure to specific social and contextual/physical cues for the behavior, including changing daily or weekly routines. Note: this may also involve 12.1, Restructuring the physical environment and/or 12.2, Restructuring the social environment; if the BCT includes analysing the behavioral problem, only code 1.2, Problem solving.</td>
</tr>
<tr>
<td>12.4. Distraction</td>
<td>Advise or arrange to use an alternative focus for attention to avoid triggers for unwanted behaviour.</td>
</tr>
<tr>
<td>12.5. Adding objects to the environment</td>
<td>Add objects to the environment in order to facilitate performance of the behavior. Note: Provision of information (e.g. written, verbal, visual) in a booklet or leaflet is insufficient. If this is accompanied by social support, also code 3.2, Social support (practical); if the environment is changed beyond the addition of objects, also code 12.1, Restructuring the physical environment.</td>
</tr>
<tr>
<td>12.6. Body changes</td>
<td>Alter body structure, functioning or support directly to facilitate behavior change.</td>
</tr>
</tbody>
</table>

### 13. Identity

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1. Identification of self as role model</td>
<td>Inform that one's own behavior may be an example to others.</td>
</tr>
<tr>
<td>13.2. Framing/reframing</td>
<td>Suggest the deliberate adoption of a perspective or new perspective on behavior (e.g. its purpose) in order to change cognitions or emotions about performing the behavior (includes 'Cognitive restructuring'); If information about consequences then code 5.1, Information about health consequences, 5.6, Information about emotional consequences or 5.3, Information about social and environmental consequences instead of 13.2, Framing/reframing.</td>
</tr>
<tr>
<td>13.3. Incompatible beliefs</td>
<td>Draw attention to discrepancies between current or past behavior and self-image, in order to create discomfort (includes 'Cognitive dissonance').</td>
</tr>
<tr>
<td>13.4. Valued self-identify</td>
<td>Advise the person to write or complete rating scales about a cherished value or personal strength as a means of affirming the person's identity as part of a behavior change strategy (includes 'Self-affirmation').</td>
</tr>
<tr>
<td>13.5. Identity associated with changed behavior</td>
<td>Advise the person to construct a new self-identity as someone who 'used to engage with the unwanted behavior'.</td>
</tr>
</tbody>
</table>

### 14. Scheduled consequences

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1. Behavior cost</td>
<td>Arrange for withdrawal of something valued if and only if an unwanted behavior is performed (includes 'Response cost'). Note if withdrawal of contingent reward code, 14.3, Remove reward.</td>
</tr>
<tr>
<td>14.2. Punishment</td>
<td>Arrange for aversive consequence contingent on the performance of the unwanted behavior.</td>
</tr>
<tr>
<td>14.3. Remove reward</td>
<td>Arrange for discontinuation of contingent reward following performance of the unwanted behavior (includes 'Extinction').</td>
</tr>
<tr>
<td>14.4. Reward approximation</td>
<td>Arrange for reward following any approximation to the target behavior, gradually rewarding only performance closer to the wanted behavior (includes 'Shaping'). Note: also code one of 59-63.</td>
</tr>
<tr>
<td>14.5. Rewarding completion</td>
<td>Build up behavior by arranging reward following final component of the behavior; gradually add the components of the behavior that occur earlier in the behavioral sequence (includes 'Backward chaining'). Note: also code one of 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward, 10.9, Self-reward; 10.10, Reward (outcome).</td>
</tr>
</tbody>
</table>
### Appendix 2

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>14.6. Situation-specific reward</strong></td>
<td>Arrange for reward following the behavior in one situation but not in another (includes 'Discrimination training'). Note: also code one of 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward; 10.9, Self-reward; 10.10, Reward (outcome).</td>
</tr>
<tr>
<td><strong>14.7. Reward incompatible behavior</strong></td>
<td>Arrange reward for responding in a manner that is incompatible with a previous response to that situation (includes 'Counter-conditioning'). Note: also code one of 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward; 10.9, Self-reward; 10.10, Reward (outcome).</td>
</tr>
<tr>
<td><strong>14.8. Reward alternative behavior</strong></td>
<td>Arrange reward for performance of an alternative to the unwanted behavior (includes 'Differential reinforcement'). Note: also code one of 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward; 10.9, Self-reward; 10.10, Reward (outcome); consider also coding 1.2, Problem solving.</td>
</tr>
<tr>
<td><strong>14.9. Reduce reward frequency</strong></td>
<td>Arrange for rewards to be made contingent on increasing duration or frequency of the behavior (includes 'Thinning'). Note: also code one of 10.2, Material reward (behavior); 10.3, Non-specific reward; 10.4, Social reward; 10.9, Self-reward; 10.10, Reward (outcome).</td>
</tr>
<tr>
<td><strong>14.10. Remove punishment</strong></td>
<td>Arrange for removal of an unpleasant consequence contingent on performance of the wanted behavior (includes 'Negative reinforcement').</td>
</tr>
<tr>
<td><strong>15. Self-belief</strong></td>
<td><strong>15.1. Verbal persuasion about capability</strong></td>
</tr>
<tr>
<td></td>
<td><strong>15.2. Mental rehearsal of successful performance</strong></td>
</tr>
<tr>
<td></td>
<td><strong>15.3. Focus on past success</strong></td>
</tr>
<tr>
<td></td>
<td><strong>15.4. Self-talk</strong></td>
</tr>
<tr>
<td><strong>16. Covert learning</strong></td>
<td><strong>16.1. Imaginary punishment</strong></td>
</tr>
<tr>
<td></td>
<td><strong>16.2. Imaginary reward</strong></td>
</tr>
<tr>
<td></td>
<td><strong>16.3. Vicarious consequences</strong></td>
</tr>
</tbody>
</table>