

Perceptions and effectiveness of episodic future thinking as digital micro-interventions based on mobile health technology

DIGITAL HEALTH
Volume 10: 1–19
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241245583
journals.sagepub.com/home/dhj



Dan Roland Persson¹ , Jakob E. Bardram²  and Per Bækgaard¹ 

Abstract

Objective: Delay discounting denotes the tendency for humans to favor short-term immediate benefits over long-term future benefits. Episodic future thinking (EFT) is an intervention that addresses this tendency by having a person mentally “pre-experience” a future event to increase the perceived value of future benefits. This study explores the feasibility of using mobile health (mHealth) technology to deliver EFT micro-interventions. Micro-interventions are small, focused interventions aiming to achieve goals while matching users’ often limited willingness or capacity to engage with interventions. We aim to explore whether EFT delivered as digital micro-interventions can reduce delay discounting, the users’ perceptions, and if there are differences between regular EFT and goal-oriented EFT (gEFT), a variant where goals are embedded into future events.

Method: A randomized study was conducted with 208 participants allocated to either gEFT, EFT, or a control group for a 21-day study.

Results: Results indicate intervention groups when combined achieved a significant reduction of $\Delta \log k = -.80$ in delay discounting ($p = .017$) compared to the control. When split into gEFT and EFT separately only the reduction of $\Delta \log k = .96$ in EFT delay discounting was significant ($p = .045$). We further explore and discuss thematic user perceptions.

Conclusions: Overall, user perceptions indicate gEFT may be slightly better for use in micro-interventions. However, perceptions also indicate that audio-based EFT micro-interventions were not always preferable to users, with findings suggesting that future EFT micro-interventions should be delivered using different forms of multimedia based on user preference and context and supported by other micro-interventions to maintain interest.

Keywords

mobile health (mHealth), micro-interventions, episodic future thinking, goal-oriented, delay discounting

Submission date: 13 September 2023; Acceptance date: 19 March 2024

Introduction

When making choices we humans do so considering a variety of factors often favoring decisions that benefit us short term and provide us with more gratification rather than less.¹ We are also interested in long-term goals such as our future health yet often we do very little to support these distant goals through our in-the-moment actions.

¹Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

²Department of Health Technology, Technical University of Denmark, Denmark

Corresponding author:

Dan Roland Persson, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, 2800, Denmark. danrp@dtu.dk



Such bias toward immediate rewards rather than potentially greater future ones is referred to as immediacy bias or as discounting of future rewards.² Previously suggested examples of discounting include: “The failing student goes out to party the night before an exam; the smoker relapses after swearing off cigarettes; the obese person eats another bag of chips.”—Odum.¹ Discounting impacts our personal health priorities by subtly shifting our focus away from future health and onto current well-being through our perception of rewards value over time. Many discrete decisions made over prolonged periods of time affect personal health, for example, if we often have a snack and relax in favor of taking a walk in the rain. However, while on paper a relatively straightforward choice in relation to long-term health, many people lean toward the more immediately gratifying choice, and especially toward those requiring less effort.¹

This discounting of future outcomes can be measured by delay discounting (DD), a behavioral economic trans-disease process¹ that measures the extent of the discounting.³ DD is a measure of the extent to which people prefer smaller immediate rewards over larger delayed ones³ and has been associated with a variety of factors affecting health.^{4–6,1,7,8} High DD has for instance been shown to be associated with overeating,⁹ lack of exercise,⁶ smoking⁴ and addiction.^{10–12} Studies have further shown DD associated with the transition of prediabetes to diabetes,⁷ poorer outcomes in Type 2 Diabetes Mellitus (T2DM)¹³ and other conditions indicative of unhealthy lifestyles.

DD is often measured through tools revolving around monetary choices.^{14,3} The classical question is; “Would you rather want \$80 dollars now or \$120 in a month?” Depending on the method employed the amounts or time frames may change between questions aiming to determine an indifference point where a subject’s opinion switches.¹⁵ The discounting rate can then be calculated as either the normalized area under the curve (AUC),^{16,17} or alternatively, as is the primary focus here, based on the correlated^{17,18} Mazur’s hyperbolic discounting model:

$$V = A/(1 + kD)$$

where V represents the subjective value, A is the actual value, D is the delay, and k is the DD coefficient.^{19,3} Lower values of AUC and higher values of k are indicative of faster discounting of future outcomes and distant outcomes being perceived as less valuable. As the interval between delays D often varies on a pseudo-exponential scale and therefore follows a skewed distribution, k is customarily log-transformed: $\log k$.¹⁷

Episodic future thinking

One way of addressing and changing a person’s DD is by using the method of episodic future thinking (EFT),² where a person projects themselves into the future through a personal, detailed,

and vivid episodic future event.^{20,21} EFT relies on a person’s ability to simulate (or “pre-experience”) events that might occur in his or her future,²⁰ such as “I am on holiday in Spain having a walk with the family enjoying the warm spring weather....” EFT is a form of prospective thinking²² thought to reduce impulsiveness by modifying the perceived value of delayed outcomes encouraging decisions with long-lasting benefits.²³ Prior research has shown that EFT is effective at reducing DD and in promoting healthy behaviors in the context of reducing the risk of T2DM,²⁴ overeating,²⁵ cigarette smoking^{4,26} and more broad substance abuse.²⁷ Research on EFT has further identified that positive future events are more effective than neutral or negatively toned events.²¹

A variation of EFT is goal-oriented EFT (gEFT), where goals are tied to the episodic future cues²⁸ in contrast to EFT.^{28,29} A general goal may for example be: “In 2 weeks I am purchasing a new computer”—O’Donnell et al.²⁸ Although goals may also be health-oriented, for example, a person might imagine that; “In a year when I am going mountain climbing again, I want to show my friends how fit I’ve become.” Here the goal is to show off one’s improved fitness level to friends through the future mountain climbing. A recent study by Athamneh et al.²⁹ further showed this adapted to health goals, finding indications that gEFT had a higher effect on intensity and elasticity in demand of cigarettes and fast-food while equally effective at reducing in-moment DD. Some studies thus cautiously suggest gEFT may be more effective at reducing the discounting rate²⁸ and may offer additional benefits through improved behaviors related to health goals.²⁹

The study of EFT is a growing area of research rooted in both cognitive neuroscience and psychology.³⁰ EFT’s ability to reduce DD has been shown in a number of contexts.^{31,20,25,4,32,9} Dassen et al. and O’Neill et al. found EFT effective in reducing discounting rate resulting in reduced snacking³² and energy intake in a public food court,²⁵ respectively. Other studies showed these effects extend to decision-making in grocery shopping³³ and demand for fast food.²³ However, these effects extend beyond diet; Stein et al. have, for example, found EFT to reduce DD directly affecting self-administration of cigarettes^{4,34} and reduced DD in persons at risk of type 2 diabetes.²⁰

A particularly promising method for delivering EFT is by mobile health (mHealth) technology, which utilizes a person’s own smartphone for the delivery of health interventions. Due to the ubiquitous availability of smartphones, mHealth technology has been suggested as very cost-effective and scalable.³⁵ A number of studies have demonstrated the use of smartphones to facilitate elements of EFT.^{36,25,37} Use of smartphone technology in EFT has, however, often been limited and mainly focused on the delivery of audio cues with little focus on design. The study by Sze et al. uses a web-based system compatible

with desktops, tablets, and smartphones,³⁶ with cue generation handled in an introductory session with a facilitator using separate open-source software and a laptop to facilitate recording.³⁶ In a follow-up study, Sze et al. further demonstrated the feasibility of using an online self-guided generation task for audio cue creation.²³ Similarly, O'Neill et al. used both participants' own smartphones and iPods to deliver audio cues, with cue generation handled through a moderated EFT cue-development interview and was recorded using a recorder.²⁵

Nevertheless, the uptake and retention of mHealth applications remain a general concern. A previous systematic review, for instance, found reported study attrition to be between 43% and 47% with indications that these may be overly optimistic.³⁸ A similar study of popular mental health applications found retention to be as low as 3.9% after just 15 days.³⁹ While the root causes behind these attrition rates are not fully understood, suggestions include lack of perceived usefulness,⁴⁰ failure to address user needs over time,⁴¹ or that patients' needs and experiences do not align with clinical judgment.⁴² A recent addition to these suggestions is that while users may be interested in digital health interventions, they may not be interested in investing a significant effort into these.⁴³

Micro-interventions

To address the current mHealth challenges, we have proposed to use EFT as "micro-interventions".⁴⁴ Micro-interventions can be defined as a highly focused type of therapeutic relying on in-moment events that employ resources to facilitate momentary changes.⁴³ Such interventions may include both just-in-time or ecological momentary interventions.⁴⁰ Given each micro-intervention focused albeit limited aims/scope, the idea of leveraging several micro-interventions has been brought forward and coined: "micro-intervention care."⁴³ This may practically be achieved through a *narrative* conveying a meaningful story through its micro-interventions and related events experienced by its users. Seeing increasing interest in recent years, current micro-intervention research primarily focuses on expanding the catalog of available micro-interventions. State-of-the-art applications of these include micro-interventions affecting stress,⁴⁵ mood,⁴⁶ and body image.⁴⁰ Howe et al.⁴⁵ delivered micro-interventions through a variety of means such as on-demand, pre-scheduled, and through a context-aware sensing system. The micro-interventions employed were named "get my mind off work," "feel calm and present," and "think through my stress" and aimed at reducing stress.⁴⁵ Meinschmidt et al.⁴⁶ allowed users a daily choice of one of four ecological momentary micro-interventions aiming to improve mood. These were viscerosensory attention (shifting attention from sensations), emotional imagery (imaging emotional situations), facial expression

(simulating emotional facial expressions), and contemplative repetition (repeating short sentences).⁴⁶

Present study

A key idea of micro-interventions is that these reduce burdens associated with getting started and the effort necessary to achieve meaningful engagement.⁴³ Given persons with high discounting already tend to discount future outcomes reducing barriers by using EFT as micro-interventions seems optimal.

While our previous study found the use of digital EFT micro-interventions was feasible, we also found that the micro-interventions were not always preferable to users.⁴⁴ Causes for this included listening to one's own voice which some participants described as uncomfortable, and a perceived disconnect between health goals and future thinking, which prompted some participants to embed health goals in their episodic futures.⁴⁴ Participants adopted this approach unprompted by study researchers and without prior introduction to gEFT with some participants moreover noting through suggested improvements that goals should be embedded in EFT.⁴⁴ However, insights offered by the study⁴⁴ were also limited by the number of participants included and the study duration with these being insufficient to determine potential effects on DD. This study, therefore, aims to investigate two research questions:

- RQ1: Can EFT micro-interventions delivered through mHealth technology reduce DD?
- RQ2: How do perceptions and effects differ between gEFT and EFT micro-interventions among users?

In this study, we investigated these research questions by modifying the previously presented mHealth application⁴⁴ to deliver both EFT and gEFT micro-interventions.

A three-arm study design was used with participants engaging in either gEFT or EFT micro-interventions with a control group receiving generic motivational materials. Participants were encouraged to engage with the app delivering micro-interventions daily for three weeks, with weekly measures of DD and motivation to engage in a chosen health behavior. The primary outcome measure is DD with secondary measures being (i) motivation to engage in healthy behaviors and (ii) user engagement in the intervention arms.

Methods

In order to investigate the effects, user perceptions, and differences between gEFT and EFT micro-interventions we opted for a three-arm study design. The study consisted of two intervention arms, that is, gEFT and EFT with the third arm serving as a control. The three-arm pilot study

was carried out following best practices of mHealth research, aiming to gain a deep understanding of how and why targeted users use or do not use the system.⁴⁷ Due to the study's technical and non-clinical objectives, the study was exempted from ethical approval by the Danish Ethical Committee (journal no. 21066249), in line with national guidelines.

EFT application

The EFT application utilized the Copenhagen Center for Health Technology Research Platform (CARP), which is an open-source research platform for digital phenotyping² and was available for both Android and iPhone operating system (iOS) smartphones. The application made use of the CARP Mobile Sensing framework⁴⁸ for data collection in compliance with the General Data Protection Regulation including in-app consent, surveys, and automatic logging of usage events.

The mobile health application was designed to deliver EFT through three types of micro-intervention events and included an app-facilitated DD assessment task³:

1. **MI1**—A self-guided generation task for creating cues.
2. **MI2**—Audio-based projection (or review) session where audio cues are experienced.
3. **MI3**—Image-based projection micro-intervention where the image represents one of the episodic futures.
4. **A1**—An assessment task aiming to measure DD through the adjusting amounts task.³

In the app-based self-guided generation task, users imagine a positive future event that they are interested in and can vividly imagine. Users are instructed through the *self-guided generation task* to imagine this event in detail including what is happening, who is present, and what they are feeling with the aim of creating a description of said event. The description is read out loud by the user and recorded through the application as “cues.” These cues are then used in the audio projection session to recall or pre-experience the recorded future events. As part of the self-guided generation task, users are also instructed to associate an image with the created future event which is used for image-based projection sessions delivered as notifications inspired by mental imagery.⁴⁹ Following our previous example about mountain climbing an image representation of this event could for instance be an image of a mountain or a previous mountain climbing trip with friends.

Users are encouraged to complete multiple self-guided generation tasks with the aim of creating seven episodic futures set at different timeframes. Users are also encouraged to engage with the applications projection sessions at least twice per day with participants able to set reminders for when to engage with these. The image-based projection sessions were sent to users twice per day during daytime as notifications randomly distributed throughout the day, but with a minimum of 2 h in between. For the purposes of this study, the EFT application was modified to automatically handle the study arms. In the following, we will present the major differences between the three study arms. A selection of screens can be seen in Figure 1.

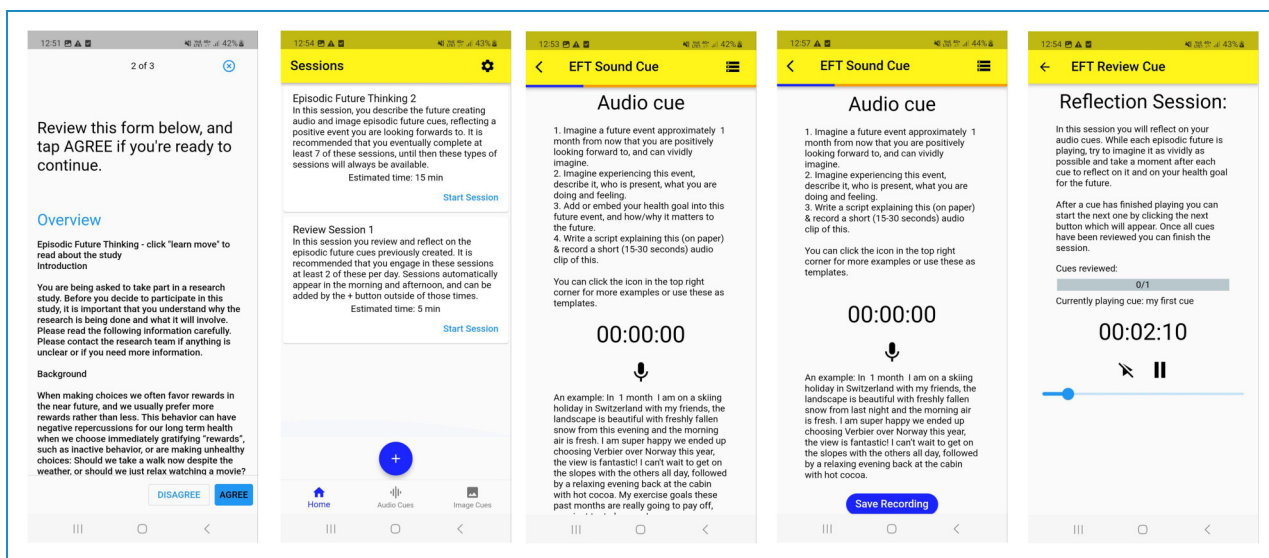


Figure 1. Design of the episodic future thinking (EFT) mobile application. The leftmost picture shows part of the integrated study consent form, with the following picture showing the application home screen of active sessions. The middle and following screens show differences in self-guided generation task instructions with the right-most photo showing the audio-based projection (or reflection) session.

EFT group. In the EFT group, users were prompted to generate episodic future events and record them for review throughout the study period. Emphasis was placed on making the episodic futures as vivid as possible. Participants were introduced to the general purpose of EFT and how it would help them achieve their health goals, but task descriptions made no direct link between the goal and episodic future events. Thus, no direct emphasis was put on relating health goals with specific episodic futures.

gEFT group. The primary differences between the gEFT and the EFT group were in the self-guided generation task where the instructions were updated to facilitate the coupling between the future and the goal. Users were asked to choose and couple a health goal to the episodic future such that emphasis is placed both on the vividness of the episodic future and the associated health goal. The image cue notifications also include a textual reminder/representation of the goal attached to the episodic future.

Control group. Participants in the control group received generic educational material on goal setting and motivation. The materials were specifically chosen to resemble materials participants would be able to find on their own using a search engine. The length of these materials was roughly equivalent to that expected of the intervention arms during the first week. Participants were instructed to choose a health goal and to apply the materials directly or use these as a basis for finding other more relevant materials in pursuit of their health goal for the remainder of the study period.

Recruitment

Participants were recruited between January 4, 2023, and January 16, 2023, at the Technical University of Denmark through online announcements, emailing lists, and announcements in classes. Interested subjects could see information about the study through a study website presenting an overall description of the study, including its purpose and aims. The website provided a link for signing up.

Measures

Demographics such as age, sex, race/ethnicity, and income were obtained using an app-based questionnaire. The primary outcome measure of the study was DD with motivation to engage in healthy behaviors measured through the treatment self-regulation questionnaire (TSRQ)⁵⁰ as a secondary outcome measure. The system usability was assessed through the post-study system usability questionnaire (PSSUQ) (version 3) with a number of other questionnaires used as experience sampling. The delivery time of measures

through the application can be seen in Table 1. In the following subsections, we will explain these measures.

DD adjusting amounts task. The DD was measured through the \$1000 adjusting amounts task³ converted approximately to the local currency, Danish Kroner (DKK), resulting in a 10.000 DKK task. In line with Epstein et al.,³ we assess delay discount at the time-frames 1 day, 1 week, 1 and 3 months, and 1, 5, 25 years. Each step of the measurement adjusts the monetary amounts by half of the previous adjustment. For example, \$500 now versus \$1000 in one month after choosing \$1000 adjusts to \$750 now versus \$1000 in one month. Each time-frame adjustment is carried out six times until the indifference point is reached. These indifference points can be used to calculate and compare the discounting rates of participants across the study.

Treatment self-regulation questionnaire (TSRQ). Motivation to engage in healthy behaviors is measured through the generic version of the TSRQ.⁵⁰ Depending on the user's chosen goals, a TSRQ question might look like "The reason I would _____ is: Because I feel that I want to take responsibility for my own health," with the dotted lines representing, for example, exercise. TSRQ is a measure capable of assessing the "self-determination continuum of motivation" for various health behaviors.⁵¹ The questionnaire's 15 questions cover a variety of factors including controlled and autonomous motivation.

Post-study system usability questionnaire (PSSUQ). The PSSUQ was used to measure the overall usability of the app.⁵² In order to provide depth and context to the answers to the PSSUQ survey, we further added a number of qualitative questions inquiring about the usability of self-guided generation task instructions and the usability of completing different sessions. Additional

Table 1. Overview of measures and survey timing.

Day	Demographic	DD	TSRQ	Other	PSSUQ
1	x	x	x	x*	
7		x	x	x*	
14		x	x	x*	
21		x	x	x*	x†

DD: delay discounting; TSRQ: treatment self-regulation questionnaire; PSSUQ: post-study system usability questionnaire; EFT: episodic future thinking.

A * denotes a step that is different for the control group (e.g. the control group is asked different questions compared to EFT conditions). A †denotes a step not carried out by the control group.

questions were also added to the end-of-study PSSUQ survey exploring the perceived good and bad aspects of the system/experience and ideas for improvements. Given the minuscule app interactions (i.e. links to materials and questionnaires) intended for the control group, PSSUQ data was not collected for this group.

Qualitative surveys. In order to explore the experience of engaging with different study conditions and the differences between these, weekly surveys were added. These surveys aimed to collect qualitative data on the experiences and provide insights into the completion of different tasks. Given the relatively low intensity of micro-interventions on a daily basis, we employed a lightweight weekly experience sampling. Following the designed timing, participants were intended to create a number of goal-oriented or EFT cues by the end of week 1. The survey thus explored experiences with self-guided generation tasks and the initial perceptions of engaging with and reflecting on different cues. Subsequent surveys additionally inquired about potential changes in the perception of activities and decision-making.

Procedures

Upon signing up for the study, participants were randomized into either of the intervention groups (goal-oriented or EFT) or to the control group. Randomization was handled through block randomization with blocks of six participants using a total of 90 possible balanced distributions, handled by a Python script producing the randomized blocks. Participants were sent email instructions on where to download the app and how to sign in. Upon first sign-in, the participant can review the subject information letter also presented through the sign-up process and can sign consent for participation in the study. After providing consent participants can use the app, which initially asks participants to complete a baseline DD measurement, an initial demographic survey, and a baseline TSRQ questionnaire for their desired health goal which included a number of suggestions: smoking cessation, diet, improved sleeping habits, increasing physical activity, or allowed users to fill in their own goal. After completing the initial surveys users can engage in the first self-guided generation task and, following its completion, daily reflection sessions for the remainder of the study period. Users are encouraged to engage with the cues at least twice per day and randomly receive one to two image notifications per day, which show a user-chosen image representation of an episodic future. On study days 7 and 14, the participant received another survey on their DD and motivation to engage in healthy behaviors. On the final day of the study, participants received a final DD measure, TSRQ survey, and the PSSUQ surveys through the app.

Participants in the control group were able to access study materials through the app similar to the intervention

arms but received different qualitative surveys reflecting the educational materials rather than micro-interventions. Participants of the control group were not asked to complete the PSSUQ survey, but instead were asked to reflect on the impact of the educational material.

Analysis

Data analysis was conducted using Python (v3.9.13) with Jupyter Notebook (6.4.12) and SciPy (v1.9.1) as well as R (v4.3.2) including base package stats (v4.3.2) and key packages lmerTest (v3.1-3), lmer4 (v1.1-35.1), robustlmm (v3.3-1), emmeans (v1.9.0), and mice (v3.16-0).

Baseline comparisons of continuous demographic variables and initial discounting by groups were performed using analysis of variance (ANOVA). Group differences in categorical data, that is, gender, ethnicity, and income were assessed using Pearson's chi-square test.

All adjusting-amounts task results not in line with the systematic criteria outlined by Johnson and Bickel⁵³ were discarded. The criteria were evidence of (a) consistent effects of sequential delays, where no indifference point exceeds the previous with more than 20% of the larger later reward, and (b) an overall reduction in discounting as a function of delays, where the last indifference must be lower than that of the first (in this case, 25 years and 1 day following Epstein et al.'s³ adjusting amounts task) by at least 10% of the larger later reward.⁵³ Following the consistency check, individual discounting rates k were calculated using Mazur's hyperbolic discounting model¹⁹ by fitting the indifference points across delays using nonlinear regression.² To account for the study design's repeated measures and to allow for individual baselines, a linear mixed-effects model was applied with the log-transformed discounting rate ($\log k$) as the dependent variable and with fixed effects of time and time in interaction with treatment, and a random intercept for each user, in accordance with Twisk et al.⁵⁴ A linear mixed-effects model using the Robust scoring equations estimator^{55,56} was also applied in order to check for any effect of potential outliers. To partially address data loss, we additionally did 100 rounds of imputing missing data through the multiple imputation by chained equations (MICE) method,⁵⁷ and subsequently pooled and reported the results of applying a similar linear mixed-effects model. This was done first for the gEFT and EFT groups combined versus the control group and secondly across all three groups. Additionally, the AUC, $AUC_{\log d}$, and $AUC_{or d}$ were calculated according to Myerson et al.¹⁶ and Borges et al.¹⁸ and checked for correlation with $\log k$. AUC was additionally modeled as above.

TSRQ scores were analyzed using t -tests comparing baseline measures to subsequent measures across outcomes for each group. Group differences were similarly analyzed using ANOVA. Qualitative data, including experience

sampling, was analyzed using color coding, affinity diagramming, and thematic analysis.

Results

A total of 208 people signed up for the study and were randomly allocated to one of the three study conditions as shown in Figure 2. Seventy participants were allocated to the goal-oriented group, with 69 participants, respectively, assigned to the EFT and control groups. A total of eight (gEFT), and nine (EFT), and 14 (control) participants completed the final end-of-study survey. All participants who downloaded the app and provided consent have been included in the analysis.

Participant characteristics

A total of 175 persons completed the demographic questionnaire with a majority of participants being 61.7% male, 38.3% female, and 34.3% of participants self-identifying as

part of a minority group. Participants were aged between 20 and 30 and all participants have or were in the process of acquiring a higher education (BSc/MSc). Participant characteristics for each group can be seen in Table 2.

Analysis revealed no significant differences between the groups' baseline demographic characteristics: gender, age, ethnicity, and income. Similarly, we observed no significant differences between baseline discounting rates. However, we did observe one statistically significant difference in baseline TSRQ scores to question 12 "Because it is easier to do what I am told than think about it" between groups ($p = .01$, uncorrected for multiple comparisons); the mean answer per condition was 3.34 (gEFT) versus 2.26 (EFT) versus 2.41 (control). It is unknown what caused this initial difference and it was not observed in subsequent measures.

Primary measures

Figure 3 shows the mean log-transformed DD rate $\log k$ per group over weeks for participants that completed the DD

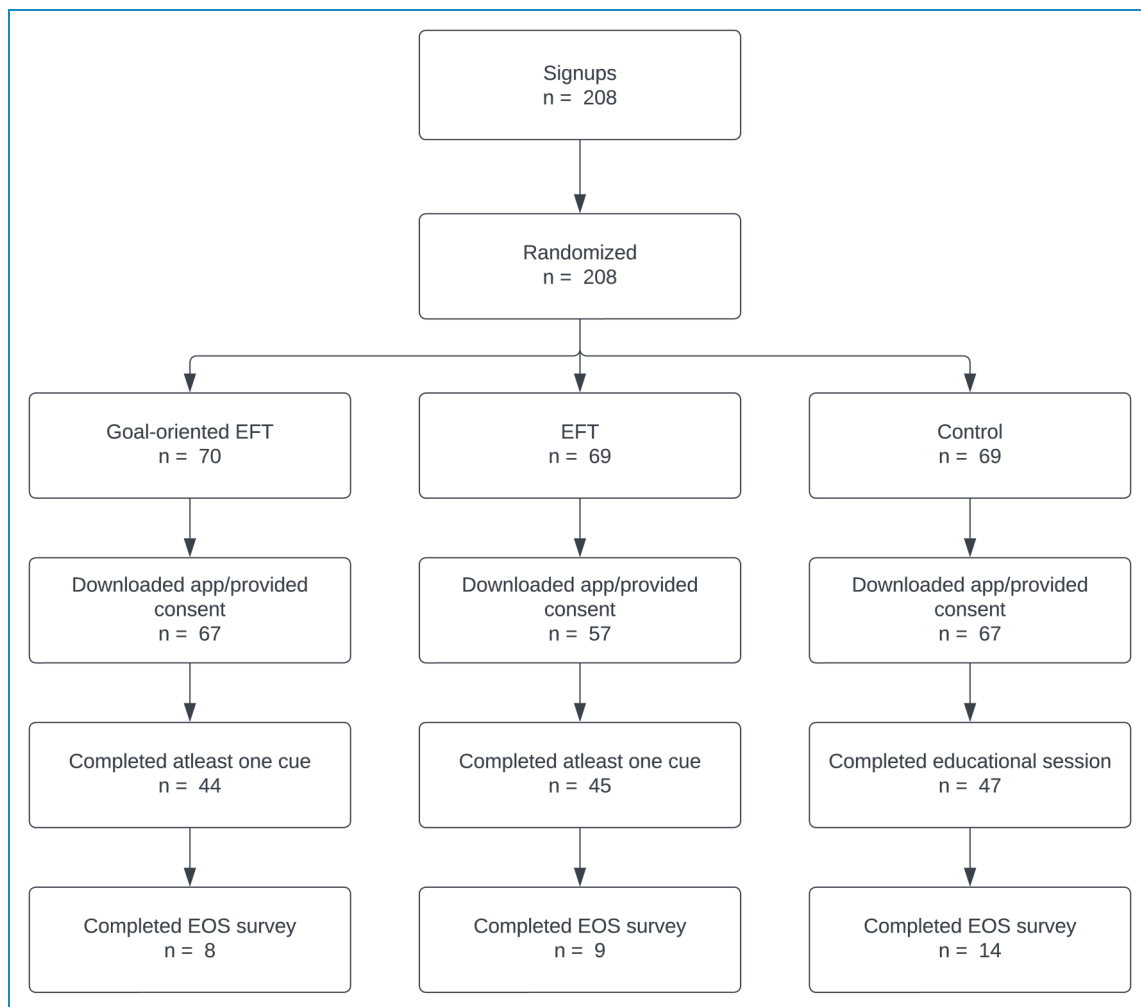


Figure 2. Study flowchart, all participants that provided consent have been included in the analysis.

Table 2. Baseline characteristics of the two intervention groups versus the control group for participants responding to the initial demographic survey.

Characteristics	gEFT	EFT	Control
<i>n</i>	59	53	63
Age mean (SD)	25.22 (4.32)	24.66 (3.26)	24.41 (4.05)
Gender			
% Male (<i>n</i>)	66.1 (39)	56.6 (30)	61.9 (39)
% Female (<i>n</i>)	32.2 (19)	41.5 (22)	36.5 (23)
% Prefer not to say	1.7 (1)	1.9 (1)	1.6 (1)
Race/ethnicity			
% Non-minority (<i>n</i>)	62.7 (37)	60.4 (32)	73.0 (46)
% Minority (<i>n</i>)	30.5 (18)	33 (18)	25.4 (16)
% Prefer not to say (<i>n</i>)	6.8 (4)	5.7 (3)	1.6 (1)
Household income			
% <190.000 DKK (<i>n</i>)	50.8 (30)	45.3 (24)	49.2 (31)
% >190.000 DKK (<i>n</i>)	20.4 (12)	26.4 (14)	34.9 (22)
% Prefer not to say (<i>n</i>)	28.8 (17)	28.3 (15)	15.9 (10)

EFT: episodic future thinking; gEFT: goal-oriented EFT; DKK: Danish Kroner.

task at the end of the specified week. By the end of the study (week 3) both the gEFT and EFT groups achieved overall reductions in mean discounting rates.

First comparing control to the two EFT conditions combined (gEFT + EFT), a linear mixed-effects model of $\log k$ with fixed effects in time (week) and time interacting with treatment (either control or gEFT + EFT) and including a random intercept for each user was significantly different ($\alpha = .05$) to a null-model using only random intercepts for each user ($\chi^2(2) = 9.29, p = .0096$) over the 203 adjusting-amounts tasks from 92 users remaining after discarding 56 inconsistent measurements according to the criteria outlined above. The overall intercept and weekly slope were $\log k = -6.472$ (SE = .172; $t(129) = -37.6, p < .001$) and $\Delta \log k = 0.052/\text{week}$ (SE = 0.090; not significant), respectively. The gEFT + EFT treatment had a weekly slope of $\Delta \log k = -0.268/\text{week}$ (SE = 0.110; $t(174) = -2.43, p = .016$). Comparing the marginal means of the gEFT + EFT group versus the control group correspondingly shows a significant difference of $\Delta \log k = -0.267$ (SE = 0.110; $t(176) = 2.42, p = .017$). Imputing

missing data using MICE resulted in 368 measurements for the 92 users, and pooling the results from 100 rounds of applying the same linear mixed-effects model showed corresponding results: The overall intercept and weekly slope were $\log k = -6.464$ (SE = 0.178; $t(294) = -36.2, p < .001$) and $\Delta \log k = 0.023/\text{week}$ (SE = 0.088; not significant vs. $\alpha = .05$), respectively. The gEFT + EFT condition had a significant weekly slope of $\Delta \log k = -0.234/\text{week}$ (SE = 0.109; $t(101) = -2.15, p = .034$).

Next, discriminating between gEFT and EFT, a linear mixed-effects model of $\log k$ with fixed effects in time (week) and time interacting with treatment (control, gEFT or EFT) and including a random intercept for each user was significantly different to a null-model using only random intercepts for each user ($\chi^2(3) = 9.81, p = .020$). The overall intercept and weekly slope were $\log k = -6.469$ (SE = 0.172; $t(129) = -37.6, p < .001$) and $\Delta \log k = 0.052/\text{week}$ (SE = 0.090; not significant vs. $\alpha = .05$), respectively. The EFT condition had a weekly slope of $\Delta \log k = -0.321/\text{week}$ (SE = 0.132; $t(172) = -2.42, p = .017$) and the gEFT condition a weekly slope of $\Delta \log k = -0.221/\text{week}$ (SE = 0.128; $t(173) = -1.73, p = .085$) without correcting for multiple comparisons. Comparing the marginal means pairwise between the three conditions correspondingly shows the difference between the EFT and the control group to be different at $\Delta \log k = -0.320/\text{week}$ (SE = 0.133; $t(173) = 2.40, p = .045$) and between gEFT and the control group $\Delta \log k = -0.221/\text{week}$ (SE = 0.139; not significant) using Tukey's method for comparing a family of estimates with Kenward-Rogers⁵⁸ approach for estimating the degrees of freedom. Imputing missing data using MICE and pooling the results from 100 rounds of applying the same linear mixed-effects model show corresponding results: The overall intercept and weekly slope were $\log k = -6.464$ (SE = 0.174; $t(317) = -37.2, p < .001$) and $\Delta \log k = 0.021/\text{week}$ (SE = 0.073; not significant vs. $\alpha = .05$), respectively. The EFT condition had a weekly slope of $\Delta \log k = -0.270/\text{week}$ (SE = 0.112; $t(137) = -2.42, p = .017$) and the gEFT condition a weekly slope of $\Delta \log k = -0.190/\text{week}$ (SE = 0.120; not significant).

Peasons correlation between $\log k$ versus AUC, $AUC_{\log d}$ and $AUC_{or d}$ were $-.89, -.99, \text{ and } -.98$, respectively. A summary of both the $\log k$ and the corresponding AUC models (not otherwise detailed) can be found in Table 3, for both conventional LME models as detailed above, with the robust scoring equations estimator (*rLMM*) and with imputation (*MICE*). The three different models show similar results with only small differences between corresponding effect sizes, well within the respective standard errors. Across all models for both $\log k$ and AUC the effect of gEFT + EFT combined is significant. So is the effect of EFT in the case of two separate conditions. The residuals of the conventional linear mixed-effects

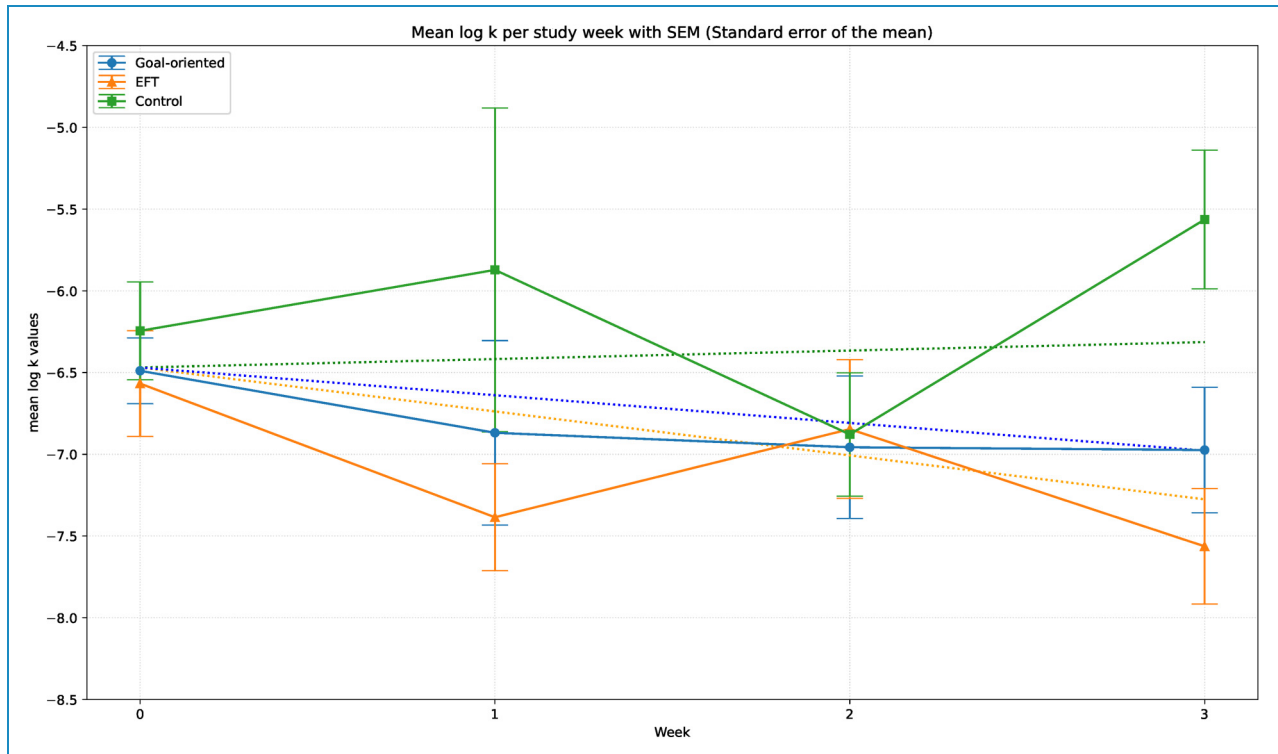


Figure 3. Mean logarithmic discounting rate ($\log k$) per group throughout the study weeks. The solid lines show the actual data and the dashed lines show the trendline derived from the LME model. Both gEFT and EFT show a declining trend over time, compared to an increasing but almost flat trend for the control group. The error bars on the solid lines indicate the standard error of the means. LME: linear mixed-effects; EFT: episodic future thinking; gEFT: goal-oriented EFT;

models were symmetric and approximated a normal distribution in quantile–quantile plots although with some outliers. However, as is visible when comparing to using the robust scoring estimator this has only a small impact on the overall model results (more than 86% of the data points had weights close to 1 and only 6%, corresponding to 13 data points, had weights $<.78$).

Table 4 shows the mean TSRQ scores per group for all participants who completed the questionnaire at the end of the specified week. While we did observe some statistically significant in uncorrected p -values to changes in TSRQ answers between weeks relative to the baseline measures (marked in bold) these were not persistent from one week to the following and are likely the result of multiple comparisons and the overall attrition in the study.

Participant flow and retention

Figure 4 shows the attrition rate and daily active users. Attrition is calculated based on the last recorded interaction with the app. An active user is one completing one or more projection sessions or self-guided generation tasks per day. The pooled attrition rate throughout the study was 88% for gEFT, 84% for EFT, and 79% for the control group. Analysis of individual usage rates further indicates that

few users accomplished the recommended session rates consistently throughout the study period, with some users completing a larger number of sessions one day while on other days completing fewer or none. Data on the precise timing of image cue notifications could not be analyzed due to a technical issue between the notification system and the mobile sensing framework.

Usability

The PSSUQ scores of the gEFT and EFT groups can be seen in Figure 5(a) and (b), respectively. They are overall similar but show a slightly higher (mean value: 2.71 vs. 3.14) overall ease of use satisfaction with the goal-oriented version of the system (PSSUQ-1: “Overall, I am satisfied with how easy it is to use this system”), even though the overall satisfaction with the system is similar (PSSUQ-16: “Overall, I’m satisfied with this system”). Participants’ answers in the gEFT group were generally more negative (mean value: 4.71 vs. 3.57) in the question PSSUQ-6: “I believe I could become productive quickly using this system” whereas those in regular EFT were more evenly divided. Experience sampling suggests this to have been caused by the perception that the system did very little to directly support the goals embedded in gEFT and thus

Table 3. The estimated intercept $\log k$ and effect sizes $\Delta \log k$ (left) as well as the corresponding AUC and Δ AUC (right) for the two treatments combined (gEFT + EFT, top) and for each of the two treatments independently (EFT and gEFT, bottom), resulting from LME models without imputation, rLMM and linear mixed effects models with imputation (MICE).

	log k			AUC		
	LME	rLMM	MICE	LME	rLMM	MICE
gEFT + EFT						
Intercept	−6.472 (0.172)***	−6.514 (0.170)***	−6.464 (0.178)***	0.285 (0.022)***	0.274 (0.022)***	0.285 (0.022)***
Base effect	0.052 (0.090)	0.025 (0.081)	0.023 (0.088)	0.002(0.012)	0.001 (0.011)	0.004 (0.012)
gEFT+EFT effect	−0.268 (0.110)*	−0.250 (0.100)*	−0.234 (0.109)*	0.039 (0.015)**	0.041 (0.013)**	0.034 (0.015)*
gEFT and EFT						
Intercept	−6.469 (0.172)***	−6.508 (0.170)***	−6.464 (0.174)***	0.285 (0.022)***	0.274 (0.022)***	0.286 (0.022)***
Base effect	0.052 (0.090)	0.024 (0.081)	0.021 (0.074)	0.002 (0.012)	0.001 (0.011)	0.004 (0.011)
EFT effect	−0.321 (0.132)*	−0.319 (0.119)**	−0.270 (0.112)*	0.046 (0.018)*	0.050 (0.017)**	0.041 (0.016)*
gEFT effect	−0.221 (0.128)·	−0.193 (0.115)·	−0.190 (0.120)	0.033 (0.017)·	0.033 (0.016)*	0.032 (0.016)*

AUC: area under the curve; EFT: episodic future thinking; gEFT: goal-oriented EFT; LME: linear mixed-effects; rLMM: LMM models with the Robust scoring equations estimator; MICE: multiple imputation by chained equations. The standard error is shown in parentheses. The uncorrected significance values are indicated with a *** for $p < .001$, ** for $p < .01$, * for $p < .05$, and a · for $p < .$; only the second model that separates gEFT and EFT needs correction for multiple family comparisons (not shown here).

that the overall system was not conducive in fulfilling these goals. However, experience sampling also indicates that gEFT participants generally had an increased goal awareness and a reduced number of statements indicating skepticism toward future thinking.

Participants in both intervention conditions were also mostly satisfied with the simplicity of using the system (PSSUQ-2), were able to complete tasks effectively (PSSUQ-3), found the system comfortable to use (PSSUQ-4), easy to learn (PSSUQ-5), and found the information provided clear (PSSUQ-9). Additionally being satisfied with instructions provided in the completion of tasks and scenarios (PSSUQ-11), and the organization of information (PSSUQ-12).

The participants were more divided on the error messages (PSSUQ-7), ease of recovering from errors (PSSUQ-8), and ease of finding information (PSSUQ-10). They were more negative toward the interface (PSSUQ-13), the use of the interface (PSSUQ 14), and the capabilities of the system (PSSUQ-15).

Qualitative data

Based on the qualitative data we identified two domains and nine key themes. The first domain relates to the micro-interventions themselves while the second relates to the facilitating application.

Micro-interventions. Task description and duration:

Participants were generally satisfied with the task instructions provided with the app and expressed satisfaction with the duration of different micro-interventions, for example, “I was surprised at how quick it was” and “... you could do your things [tasks] relatively quickly.” The self-guided generation task was generally perceived as more difficult than subsequent reflection tasks, with tasks involving faraway futures being the most difficult to imagine.

Statements made by the participants allude to a potential learning curve effect in creating good episodic futures, with some participants noting that they became unhappy with their initial episodic future after a few review sessions, prompting them to remake or improve the cues. Others noted that it became easier to imagine and create futures over time, especially those closer to the present. While most participants found the provided examples useful and the instructions sufficient for task completion, some users still noted difficulties getting started on their first cue.

Listening to one’s own voice: A significant number of participants reported discomfort from listening to their own voice. This is reflected in frequent comments such as “I also cringed a little at listening to my own voice trying to motivate me” and “I don’t think anyone likes hearing their own voice.” Surprisingly, this caused some users to use the

Table 4. Means of TSRQ answers throughout the study per condition.

	Baseline			Week one			Week two			Week three		
	GO	EFT	CTL	GO	EFT	CTL	GO	EFT	CTL	GO	EFT	CTL
Q1	5.76	5.42	5.57	6.0	6.10	6.67	6.16	5.91	5.85	6.18	4.78	6.00
Q2	3.34	3.26	3.35	3.5	3.35	3.67	3.63	3.23	3.35	3.91	3.89	3.31
Q3	5.97	6.16	5.89	5.8	6.30	6.00	6.26	6.18	5.80	6.00	4.89*	6.23
Q4	2.38	1.84	2.22	2.4	1.85	1.67	2.53	2.27	2.15	2.18	2.33	2.46
Q5	2.59	2.48	2.62	2.2	2.35	3.00	2.47	2.59	2.70	2.09	2.33	2.15
Q6	5.62	5.52	5.22	5.0	5.55	6.00	5.95	5.59	5.15	5.36	4.78	5.38
Q7	4.38	4.00	4.14	3.6	4.15	5.00	4.47	4.18	3.95	4.27	4.00	3.77
Q8	5.34	5.10	5.19	5.1	4.90	6.00	5.95	5.09	4.50	5.36	4.89	5.08
Q9	2.28	1.81	2.38	2.4	2.10	2.33	2.47	1.91	2.30	1.73	2.11	2.46
Q10	2.76	2.65	2.78	3.2	2.30	4.00	2.68	2.50	2.50	2.36	3.22	2.62
Q11	5.34	5.48	4.78	5.4	5.70	5.33	6.37**	5.36	5.05	5.82	5.67	5.38
Q12	3.34	2.26	2.41	3.3	2.55	2.00	3.32	2.41	2.10	2.55	2.78	1.92
Q13	5.76	5.77	5.62	5.6	5.90	5.00	5.47	5.73	5.75	5.55	4.67	5.77
Q14	3.66	2.55	2.81	2.5	2.75	3.00	4.16	2.68	2.55	3.82	2.89	2.46
Q15	1.72	1.84	2.11	1.8	2.35	2.67	1.89	1.91	2.35	1.55	2.11	2.23

TSRQ: treatment self-regulation questionnaire; GO: goal-oriented; EFT: episodic future thinking; CTL: control.

A · indicate the difference to the baseline for the same condition is statistically different at $p < .1$. A * denotes $p < .05$ and ** $p < .01$. Note that the marked p-values have not been corrected for multiple comparisons, and are not been statistically significant when this is done.

system in unintended ways: “I skipped the audio cue because I don’t like the sound of my own voice,” with the user instead relying solely on image cues. While this topic gradually became less frequent in the latter half of the study, some participants continued to note annoyance with listening to their own voices, for example, “It became a bit less painful to listen to my audio recordings, but I still wouldn’t say I’m comfortable with it.” This is also reflected in the participants’ various suggestions on how the system could be improved, for example, “I don’t like using voice recordings to reflect on future plans. A text-based to-do list works much better.”

Contextual challenges: Participants also noted various contextual challenges to engage with EFT audio cues in everyday life, for example, “I searched for a quiet space and listened to my audio cue” and “I tried to be in some quiet surroundings that gave me space to reflect,” noting that this could occasionally be a challenge in everyday

life. This issue is especially profound when attempting to complete the self-guided generation task, as there is a need for a quiet place to record it, and relative to the reflection sessions a need for more time to consider/create a vivid episodic future.

Notifications: The amount and frequency of notifications were perceived as appropriate by the users, with the relatively low intensity and volume of notifications being cited as positive. Whereas the participants frequently mentioned barriers to the usage of audio cues they noted satisfaction with the image cues. Some users further perceived the reminders for audio-based reflection sessions as being enough to prompt them to start reflecting on the future even before engaging with the audio cues: “the notification of the reflect session was enough to get me thinking about my future cues.”

Perceived effects: When prompted to reflect on the effects of the interventions both intervention conditions

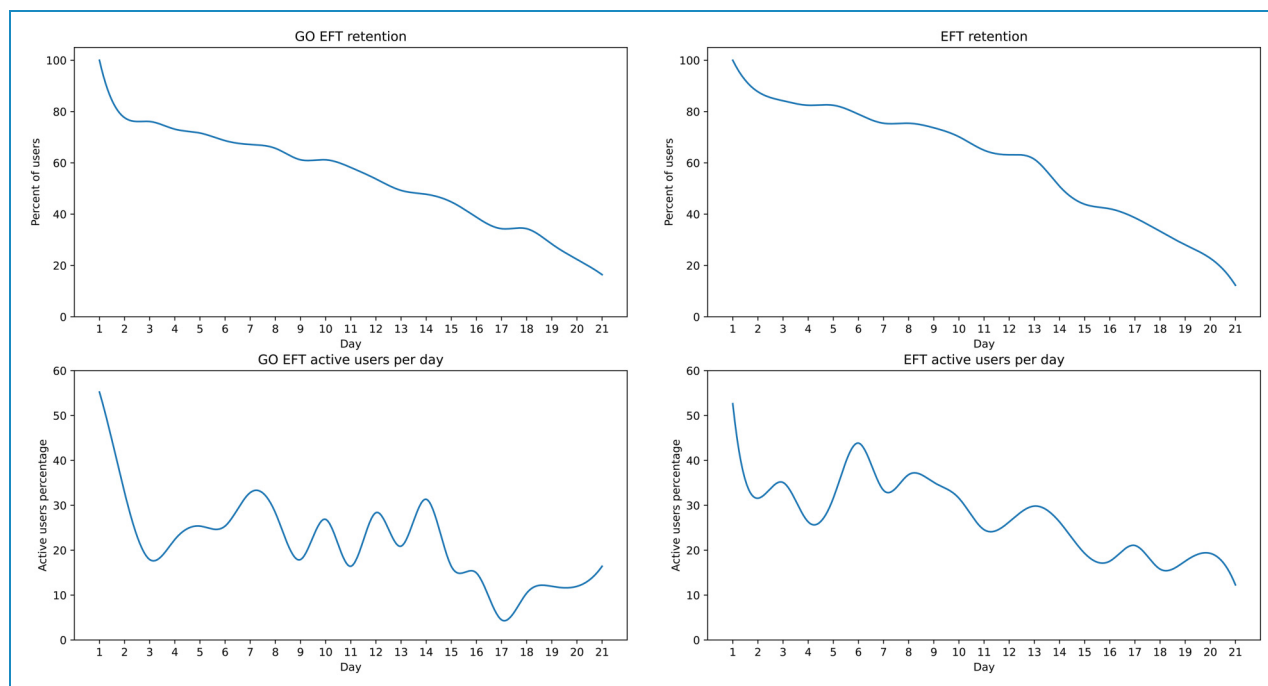


Figure 4. The two topmost plots show attrition for the gEFT (left) and EFT (right) in percent of the total number of users signed up, across days counting from when the individual participant was signed up. The bottom two plots show the number of active users on each day counting from the sign-up (gEFT left and EFT right). EFT: episodic future thinking; gEFT: goal-oriented EFT.

noted they had not perceived any changes to their decision making. However, some participants in the goal-oriented group also noted that they were more aware of their health goals in everyday life and that they thought about the future more than prior to enrolling in the study, with, for example, one participant noting: “it did motivate me to go on a run one day when I wasn’t feeling it.” The image cues were likewise noted as a positive reminder of one’s cues: “I have found it fruitful to do so [...]. I especially like the image cues that remind me that the future does exist, even when I am caught up in the present.” Participants in the EFT group similarly noted they were more aware of the future with one participant adding the increased awareness had made him more motivated to exercise. Interestingly, some participants in the EFT group seemingly adopted a “goal-oriented” approach to EFT unprompted by task instructions, for example, “I used it [EFT] as a reminder of what I needed to do, to get to the future I wanted.”

Participants in the goal-oriented group also noted: “it was really nice, I did not expect that listening to my future goals would give me a positive perception” and “listen to it and considered whether I finish the aims,” whereas participants of the EFT group were generally more skeptical about how considering the future would help them, for example, “I couldn’t find the motivation in the recording.” Some participants also cited the lack of noticeable effects as discouraging for usage: “The cues

don’t do anything for me. I don’t think my brain works like that.”

Perceived usefulness: When comparing participants’ initial experiences, we see a number of users in the EFT group express confusion, doubt, or skepticism toward engaging in future thinking, whereas these expressions are almost entirely absent from those of the goal-oriented group.

Some participants through their experiences also described a perceived preference against the EFT micro-interventions. Participants, for example, described how they were already working with various health goals and therefore perceive future thinking (both gEFT and EFT) micro-interventions as adding very little value. Others also noted that future thinking added very little as they were already often aware of, focused on, or reflecting on the future. A few participants speculated that the interventions did not suit their tastes. Some participants cited various factors as affecting perceived usefulness such as the lack of feedback or highlighting the consequences, for example, in relation to discounting the future, and the lack of concrete steps in achieving certain episodic futures.

Users in both intervention conditions found the experience of engaging with future thinking positive: “I think it brings value and meaning to the actions you do in the present” (gEFT) and “I have found it fruitful to do so. I especially like the image cues that remind me that the

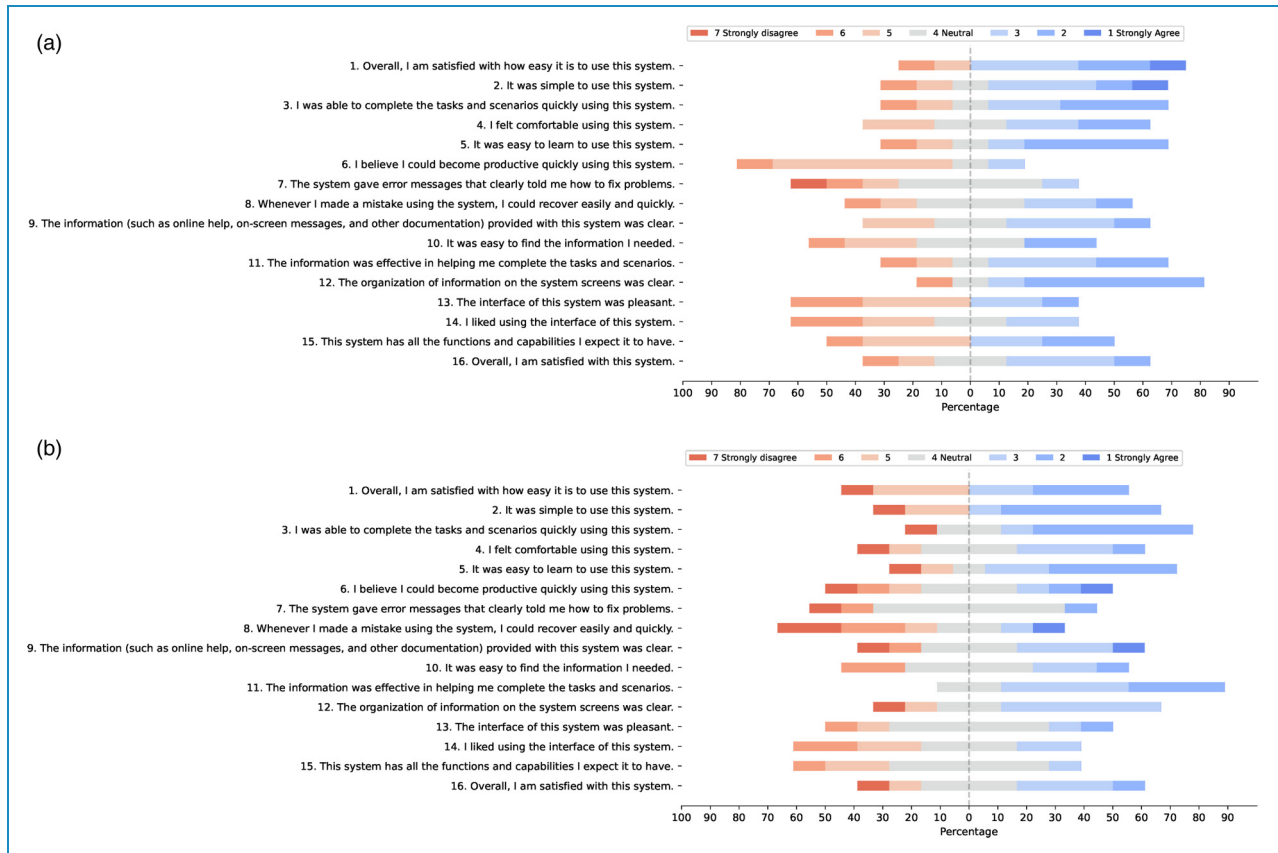


Figure 5. Diverging stacked bar chart of the results from PSSUQ by the type of future thinking served showing the percentage of respondents answers based on the Likert scale (color coded). (a) gEFT PSSUQ scores; (b) EFT PSSUQ scores. PSSUQ: post-study system usability questionnaire; EFT: episodic future thinking; gEFT: goal-oriented EFT.

future does exist, even when I am caught up in the present” (EFT).

Experience over time: Participants noted some annoyance with the recordings over time, citing them to become somewhat boring and repetitive. Some participants also noted changing perceptions toward the intervention over time. “I was more invested in the start. As they didn’t do anything for me after a few weeks, I lost interest.”

During the course of the study, participants also noted that some of their short-term episodic futures came to pass and that they subsequently created new ones keeping them in the future.

mHealth application. Interface: Perceptions of the application’s interface were generally varied in line with the qualitative questions reflected in PSSUQ-14 and PSSUQ-15. Several participants, for example, appreciated the “Simplicity, ease of use and notifications” and that the system was “not overwhelmed with information.” However, some participants also added they were not satisfied with the color scheme of the application, the design of certain screens, and the general polish of the interface.

Bugs: Participants also reported some software bugs specifically relating to background processes on both iOS and Android as the mobile sensing and notification systems were initially subject to termination by the operating system on certain versions. Similarly, these issues caused some frustrations on iOS in relation to self-guided generation tasks, for instance, if the users “left” the task unfinished with the app in the background it could sometimes force users to rerecord their cue. Some users also commented on the power consumption of the application noting this to be a negative of usage, likely also caused by mobile sensing.

Suggestions for improvements: Participants themselves provided several suggestions for improving the usability of certain areas of the system, such as improving the flow of the cue-creation by splitting the task into smaller steps and allowing more flexibility and other types of media used as part of the intervention. In both intervention conditions, participants were expected to engage with both audio and image cues, however, user comments indicate that different forms of media are preferable depending on the context, that is, it can be difficult to find a quiet place to listen to audio thus making

text/images preferable contextually. Furthermore, certain types of media may be preferable to different users as indicated by reported discomforts related to listening to one's own voice.

Discussion

Principal results

Returning to our research questions regarding the effectiveness of EFT-based micro-intervention delivered through mHealth and the perceptions and effectual differences between gEFT and EFT, this study adds to the growing literature on the effects of EFT by showing that gEFT and EFT used for a limited period as micro-interventions may be capable of modifying DD.

The present study demonstrated a modest but statistically significant weekly effect of $\Delta \log k$ at $-.27$ (LME model), corresponding to $-.80$ over the three weeks, when analyzing both treatment groups gEFT and EFT groups in combination versus the control group.

Modeling $\log k$ using rLMM or MICE shows very comparable results within the estimated confidence intervals, although we note that a relatively large amount of data was imputed for the MICE model. Similarly, modeling ΔAUC using LME, rLMM, or MICE shows a statistically significant weekly effect of $.039$ (LME model), corresponding to $.117$ over the three weeks. The significance of the effect thus appears robust across the three models and the two target variables.

Corresponding effects were observed when modeling each condition independently, although at different significance levels; only the effect of EFT was statistically significant. The effect on $\Delta \log k$ of EFT was $-.32$ (LME model) or $-.96$ over the three weeks, whereas the effect of gEFT (not significant) was $-.22$ weekly or $-.66$ over the three weeks. Again, the rLMM and MICE models showed very similar results.

The difference between the statistical significance of the effect of the combined treatment versus each treatment individually is likely related to the loss of statistical power caused by the overall attrition in the study.

All but six of the participants (four in the control group and two in the EFT group) for whom we had usable data had an initial $\text{AUC}_{\log d}$ higher than $.50$. This may hint that few of the participants would benefit strongly from an intervention targeting an improvement in DD. However, with so few participants potentially needing an intervention, it becomes unfeasible to model if they benefit more than the rest. Excluding participants with an initial $\text{AUC}_{\log d}$ higher than a cut-off at $.75$ does not significantly change the effect size. At lower cut-off points, the statistical power of the model starts to greatly diminish.

Nevertheless, relative to a recent six-month trial with an effective reduction in discounting rate of -1.99 ($\log k$ mean

difference),²⁴ our results are modest, although it was achieved in just three weeks. Our findings suggest changes can be achieved over short time periods. These results are interesting for two reasons: first, the modification of the discounting rate occurred despite relatively low adherence compared to the recommended frequency of interacting with audio-based reflection sessions. Secondly, these changes occurred relative to the low intensity of the interventions employed, with image cues being consumed in a matter of seconds and audio-based reflection sessions in a few minutes throughout the day. However, despite these improvements in discounting rate, similar improvements were not found in motivation to engage in chosen health behaviors. Nevertheless, the results from the experience sampling (see the "Perceived effects" section) suggest some users did perceive changing in-moment behaviors due to the micro-interventions. While relatively few users reported such experiences these findings tentatively support previous EFT research on modifying in-moment behaviors.^{25,33} Our results are also in contrast to Voss et al.'s similarly scoped one-month study, which despite observing behavioral modifications did not observe changes in DD.⁵⁹ However, a key difference between our study and that by Voss et al. was our inclusion of mental imagery micro-interventions to support EFT micro-interventions.

While we had expected gEFT to outperform EFT in terms of user retention due to the more direct tie-in to health goals, this was not observed in the pooled attrition. Nevertheless, the differences in reported experiences from the first week do indicate a clearer perceived value from the inclusion of goals relative to EFT: participants in the EFT group expressed concerns about the effectiveness of the intervention whereas such comments were by large absent in the gEFT group. These indications support previous findings that there may be advantages to gEFT over EFT,²⁹ but also hints that there may be small (in this case also not significant) differences in reduction of DD for gEFT.²⁸

User perceptions

These positive results are, however, contrasted by a significant attrition rate across all study conditions, with the control group seeing the lowest overall attrition. Relative to, for example, Epstein et al.'s six-month trial focused on persons with prediabetes,²⁴ ours employs a generic audience without known preexisting conditions. We speculate this difference in users and their motivation to change may partly explain our higher attrition.

Paradoxically this could also be due to a "chicken or the egg" type of problem, in that DD prevents people from engaging with interventions aiming to modify DD. On a user experience level, these findings emphasize the paramount importance of strong and clearly perceived values from apps and micro-interventions when attempting to

leverage future thinking. Previous formative usability testing indicates there may be a potential in highlighting/coupling future thinking to the potential consequences, such as diabetes or in general the health impact.⁴⁴ However, we did not observe any correlation between initial DD and how long participants stayed in the study (Pearson's $R = -.052$, $p = .65$).

We also speculate that the attrition seen may in part be caused by our study design, specifically in its resemblance to "in the wild" adaptation of mHealth. In line with other real-world app usage, the sign-up process, app download, and setup process could be completed in a short amount of time on the app and with little effort, indicative of the low burden generally associated with real-world app installation.³⁹ However, the self-guided generation task requires relatively higher efforts and considerations on the users' part, as was also reflected by the reported experiences. Consequently, we see indications supporting the idea that some participants may have joined the study out of curiosity rather than genuine interest or desire for behavior change.⁴⁰ Overall, these observations may support Baumel et al.'s point that while people are interested in digital interventions they may not be interested in investing much effort into these interventions,⁴³ reflected in a high sign-up and download rate compared to actual engagement with the interventions. Nevertheless, it can also be argued that our recruitment criteria did not reflect a group with strong discounting or a particular need for EFT micro-interventions which may have also contributed to observed attrition and lack of perceived usefulness. Future works could investigate whether these user perceptions persist among groups with a more explicit need for EFT.

While comments about listening to one's own voice were generally less prominent in the latter half of the study, it remained a prevalent point of frustration. A potential reason for this could be that participants with more resilience toward listening to their own voice were retained while those unable to overcome this barrier were lost due to attrition. Regardless, the quantity and negative connotations from experiences sampled within the first week indicate this is a significant barrier to future thinking in the wild. Participants' own suggestions show a preference for reflecting on the future using other multimedia formats, such as the image cues included in the study. One concrete suggestion included creating futures as text, utilized similarly to image cues through notifications. The preferences for image-based reflections also support Chan et al.'s⁴⁹ suggestion that mental imagery may be a convenient low barrier technique to promote healthy behaviors. A similar length study with an equivalent age group, while finding positive indications of behavior change from EFT, observed no meaningful differences between pre/post discounting rate.⁵⁹ This could indicate that the use of mental imagery in addition to EFT may have enhanced the effects of EFT

or itself acted as a primary contributor to the observed changes to DD.

In contrast to prior works, experience sampling did not reveal any fatigue from notifications themselves,^{60,61} but did see a need for novelty toward end-of-study, in line with prior works on micro-interventions.^{62,45} These observations are interesting as they seem to support the idea that micro-interventions by their short nature can be delivered with little burden to the individual.⁴³ Despite the relatively high volume of events (5 per day) recommended at the start of the study, experience sampling did not reveal negative perceptions about the frequency of micro-intervention events, with some participants citing the overall notification frequency as positive. In contrast, participants were more vocal about their perceptions of EFT including the contextual appropriateness of audio-based sessions and preference against sessions utilizing their own voice. Personalization options for notifications may have contributed to this positive perception as participants could change the timing of reminders (with no notification sent if a session was completed prior to the reminder), initially scheduled in line with prior studies, that is, prior to the first and last meals of the day.³⁶ However, we cannot fully discount the possibility that session frequency and notifications may have contributed to observed attrition. This also suggests future works on micro-interventions should more explicitly explore the relationship between event frequency over time and factors such as engagement, effects, and attrition.

It is further important to contextualize the attrition seen in this study with the nature of micro-interventions. We would argue that using any single micro-intervention for a prolonged period of time goes against the ethos of micro-interventions, and indeed a significant part of the potential value of micro-interventions. Micro-interventions aim to provide meaningful therapeutics within the limited confines of users' willingness to exert effort into the interventions.⁴³ Given this short-term nature, we expect micro-interventions to operate in limited windows of opportunity. To address the limited scope of each micro-intervention, several works have emphasized the potential of leveraging combinations of micro-interventions in mHealth through narrative structures.^{43,63,40} Such narrative structures would allow for micro-interventions to be deployed, switched, or abandoned based on a variety of user-centered factors detected by the mHealth system, for example, changing needs,⁴¹ preference,⁶⁴ the need for novelty,⁴⁵ and various barriers to use.⁶⁵ We see several of these factors reflected in our qualitative feedback with, for example, some participants noting the EFT micro-interventions alone were not enough to motivate them or that they gradually became bored of the interventions, suggesting a need for novelty. This is in line with Paredes et al.'s⁶² findings, which advocated for novelty as an important factor in avoiding boredom and subsequent attrition. We thus find our results promising in

the context that each micro-intervention should ideally only be used for a limited amount of time and in combination with other micro-interventions through an overall narrative structure. EFT micro-interventions present an interesting addition to the expanding catalog of micro-interventions for use in narratives, although such narrative tailoring of micro-interventions and the potential synergies between them is an emerging field of research and generally under-explored.

Several insights can also be drawn in the broader context of the micro-intervention research. Our findings support Howe et al.'s⁴⁵ observation that users overwhelmingly want intervention content that is simple, easy to do, and requires minimal effort. We see this reflected in our experience sampling wherein many users note a preference toward the image cues and comments that audio-session reminders were often enough to prompt users to think about the episodic futures. Similarly, our findings support Meinschmidt et al.'s⁴⁶ suggestion that micro-intervention sessions are generally well tolerated by users given the lack of negative feedback and indeed also positively reported perceptions of task durations.

However, in line with the findings of Fuller et al.,⁴⁰ we note that attrition of our mHealth system utilizing micro-interventions remains largely the same as that of traditional real-world mHealth interventions.³⁹ These results seem to indicate micro-interventions do not actually reduce barriers related to engagement.⁴⁰ Nevertheless, we speculate that the limited number of micro-interventions included in our app with uniform aims are not sufficient to maintain engagement. Moreover, the presented micro-interventions do not necessarily address users' actual needs or comply with their preferences both previously established factors in micro-intervention usage.^{62,40,43,45}

From a technological standpoint, our results thus indicate that mHealth systems utilizing micro-interventions should preferably include different micro-interventions and multimedia variants to facilitate engagement based on for instance context, needs, and user preferences. The current system could for instance be extended with capabilities allowing it, rather than the user to read out loud the episodic futures mitigating issues related to own's voice. Alternatively, the system could allow users themselves to read out textual representations in the moment of engagement. Based on users' perceived in-moment behavior changes and the performance of mobile sensing, we further believe it possible to extend the system with simple contextually relevant events (i.e. just-in-time). For instance, using Bluetooth Beacons to create simple context events⁶⁶ aiming to affect momentary decisions.²⁵ However, judging by user comments real-world use of context-aware mobile sensing in mHealth needs to be carefully balanced with power consumption to avoid user frustration and attrition.

Experience sampling also supports the observations from our previous study namely that certain users of EFT

adopt a goal-oriented approach unprompted.⁴⁴ While participants of the gEFT group were specifically instructed to embed goals into their episodic futures, this was not the case for the EFT group and future works could explore the frequency of and causes for participants doing so.

A strength of this work is that it looks specifically at user perceptions of gEFT and EFT as micro-interventions in a setting reminiscent of real-world usage. As previously discussed, users could access the study content relatively quickly and were not motivated by incentives or reminders from study researchers. However, an obvious downside and limitation to the present work are in terms of DD data fidelity and quality as few users were retained throughout the study. Resulting attrition across conditions reassemble the rate found by Baumel et al.³⁹ relating to real-world usage suggesting that although EFT may be a potentially beneficial intervention a significant proportion of users are not interested or motivated on their own to engage with these long term.

Overall our insights can be summarized by five design reflections for future mHealth systems wanting to utilize EFT micro-interventions. Firstly, these micro-interventions should avoid using recordings of the user's own voice and secondly, allow users different means of engaging with future thinking based on individual preference, for example, computer-generated audio, text, and imagery. Given the low intensity and short nature of micro-interventions, we find it unlikely that most users can overcome the voice barrier without significant frustration. Thirdly, EFT micro-interventions should preferably be based on gEFT given the slight advantages presented by this approach. Fourthly, the micro-interventions should be served in a more context-aware fashion aiming to engage users at opportune moments with resources matching the context and user preference. Lastly, EFT micro-interventions should not be used in isolation but rather as part of a larger narrative aiming to maintain user interest and support health goals with other micro-interventions benefiting from a reduced DD.

Limitations

Three major limitations should be noted when interpreting the results of this study. First, our participant group is homogeneous, given that all participants were recruited from an institution of higher education. Moreover, participants were recruited from interested individuals who appear not to be particularly strong discounters based on their initial discounting rate and thus do not have a particular need for reducing their DD. While we aimed to keep the micro-interventions generally usable, future work therefore should aim to target user groups that have more explicit needs or could benefit from EFT. Our results may therefore not be immediately generalizable to the larger population. Second, the significance and weight of our results are

limited given the relatively high, although not unusual, attrition observed across conditions. Future work is therefore necessary to confirm the validity of the reported results more generally. Third, we did not explore which of the two types of reflection (audio-based or image-based) elicited the largest contribution to the outcome. Future work therefore needs to explore the different effect sizes comparing audio cues versus image cues.

Although beyond the scope of the current work, future work could also explore how participants choose image representations of episodic futures and goals more explicitly. In the current work, experience sampling tentatively indicates images did successfully prompt participants to think about the episodic futures in line with mental imagery: “Mental images are cognitive constructions of hypothetical events or reconstructions of real events and that they can act as rich schema or representations of goals, actions or behaviors.”⁴⁹ However, it is possible other stimuli such as episodic memories may have been triggered by the images.

Additional work is likewise needed to explore whether or not EFT micro-interventions are effective at increasing uptake and motivation to engage in other/subsequent micro-interventions as part of a narrative. Lastly, we do not know whether the modified discounting persists past the initial micro-intervention usage period suggesting the need for future studies to include a follow-up period.

Conclusion

This paper has presented a randomized feasibility study investigating the use of mHealth technology for EFT in order to improve DD in health. The study enrolled 208 participants, who were assigned to three conditions; Classic EFT, gEFT, and a control group. With respect to RQ1 (“Can EFT micro-interventions delivered through mHealth technology reduce DD?”), our results indicate that both gEFT and EFT can work as micro-interventions and are capable of modifying discounting rates. Regarding RQ2 (“How do perceptions and effects differ between gEFT and EFT micro-interventions among users?”), we did not detect any statistically significant difference between the effects of the two intervention groups. Nevertheless, experience sampling suggests that gEFT made participants more aware of their goals in everyday life compared to those in the classic EFT group, with participants also being more positive toward future thinking itself. gEFT may thus be better aligned with the short-term nature of micro-interventions.

However, these results should be viewed in light of a large and rapid attrition rate. Factors associated with this attrition include listening to one’s own voice, lack of perceived value, and the contextual appropriateness of micro-interventions.

Bearing these limitations in mind, we conclude that the use of EFT micro-interventions is feasible with results

indicating that the micro-interventions can affect DD and in-moment behaviors in everyday use. If the attrition rate can be mitigated—for example, by better matching delivered EFT resources to users’ context and preferences through mHealth—EFT micro-interventions could play a key role in micro-intervention systems as a narrative component.

Contributorship: The study was planned by PD and BP. PD researched the literature, applied for ethical approval, designed/modified the initial application/micro-interventions, and app implementation. BJ contributed to the implementation of the application. PD and BP were responsible for participant recruitment, enrollment, and initial analysis, with PD writing the first manuscript draft. All authors were involved in the data analysis, manuscript creation, and editing. All authors reviewed and approved of the final version of the manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Ethical Approval: Due to the study’s technical and non-clinical objectives, the feasibility study was exempted from ethical approval by the Danish Ethical Committee (journal no. 21066249).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is part of the iPDM-GO project⁶⁷ that has received funding from European Institute of Innovation and Technology (EIT) Health. EIT Health is supported by the EIT, a body of the European Union that receives support from the European Union’s Horizon Europe research and innovation programme. This work is supported by the Copenhagen Center for Health Technology (CACHET), which is a strategic partnership between the Capital Health Region of Denmark, the City of Copenhagen, the University of Copenhagen, and the Technical University of Denmark.

Guarantor: BP.

ORCID iDs: Dan Roland Persson  <https://orcid.org/0000-0002-4258-6949>

Jakob E. Bardram  <https://orcid.org/0000-0003-1390-8758>

Per Bækgaard  <https://orcid.org/0000-0002-6720-1128>

Notes

1. A process occurring across a wide range of conditions making domain findings universally applicable.⁶⁸
2. <http://carp.cachet.dk/>

References

1. Odum AL. Delay discounting: I’m ak, you’re ak. *J Exp Anal Behav* 2011; 96: 427–439.

2. Bickel WK, Stein JS, Paluch RA et al. Does episodic future thinking repair immediacy bias at home and in the laboratory in patients with prediabetes? *Psychosom Med* 2020; 82: 699–707.
3. Epstein LH, Paluch RA, Stein JS et al. Delay discounting, glycemic regulation and health behaviors in adults with prediabetes. *Behavioral Med* 2021; 47: 194–204.
4. Stein JS, Wilson AG, Koffarnus MN et al. Unstuck in time: episodic future thinking reduces delay discounting and cigarette smoking. *Psychopharmacology* 2016; 233: 3771–3778.
5. Stein JS, Sze YY, Athamneh L et al. Think fast: rapid assessment of the effects of episodic future thinking on delay discounting in overweight/obese participants. *J Behav Med* 2017; 40: 832–838.
6. Tate LM, Tsai PF, Landes RD et al. Temporal discounting rates and their relation to exercise behavior in older adults. *Physiol Behav* 2015; 152: 295–299.
7. Epstein LH, Paluch RA, Stein JS et al. Role of delay discounting in predicting change in hba1c for individuals with prediabetes. *J Behav Med* 2019; 42: 851–859.
8. Epstein LH, Jimenez-Knight T, Honan AM et al. Imagine to remember: an episodic future thinking intervention to improve medication adherence in patients with type 2 diabetes. *Patient Prefer Adherence* 2022; 16: 95–104.
9. Daniel TO, Stanton CM and Epstein LH. The future is now: Comparing the effect of episodic future thinking on impulsivity in lean and obese individuals. *Appetite* 2013; 71: 120–125.
10. Amlung M, Vedelago L, Acker J et al. Steep delay discounting and addictive behavior: A meta-analysis of continuous associations. *Addiction* 2017; 112: 51–62.
11. Forster SE, Steinhauer SR, Ortiz A et al. Evaluating effects of episodic future thinking on valuation of delayed reward in cocaine use disorder: A pilot study. *Am J Drug Alcohol Abuse* 2021; 47: 199–208.
12. Snider SE, LaConte SM and Bickel WK. Episodic future thinking: Expansion of the temporal window in individuals with alcohol dependence. *Alcohol Clin Exp Res* 2016; 40: 1558–1566.
13. Campbell JA, Williams JS and Egede LE. Examining the relationship between delay discounting, delay aversion, diabetes self-care behaviors, and diabetes outcomes in us adults with type 2 diabetes. *Diabetes Care* 2021; 44: 893–900.
14. Kirby KN, Petry NM and Bickel WK. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *J Exp Psychol Gen* 1999; 128: 78.
15. Ad Matta, L Gonçalves F and L Bizarro. Delay discounting: Concepts and measures. *Psychol Neurosci* 2012; 5: 135–146.
16. Myerson J, Green L and Warusawitharana M. Area under the curve as a measure of discounting. *J Exp Anal Behav* 2001; 76: 235–243.
17. Yoon JH, De La Garza R, Newton TF et al. A comparison of Mazur's k and area under the curve for describing steep discounters. *Psychol Rec* 2017; 67: 355–363.
18. Borges AM, Kuang J, Milhorn H et al. An alternative approach to calculating area-under-the-curve (AUC) in delay discounting research. *J Exp Anal Behav* 2016; 106: 145–155.
19. Mazur JE. An adjusting procedure for studying delayed reinforcement. *Quant Anal Behav* 1987; 5: 55–73.
20. Stein JS, Craft WH, Paluch RA et al. Bleak present, bright future: II. Combined effects of episodic future thinking and scarcity on delay discounting in adults at risk for type 2 diabetes. *J Behav Med* 2021; 44: 222–230.
21. Ye Jy, Ding Qy, Cui Jf et al. A meta-analysis of the effects of episodic future thinking on delay discounting. *Q J Exp Psychol* 2021; 17470218211066282.
22. Atance CM and O'Neill DK. Episodic future thinking. *Trends Cogn Sci* 2001; 5: 533–539.
23. Sze YY, Stein JS, Bickel WK et al. Bleak present, bright future: Online episodic future thinking, scarcity, delay discounting, and food demand. *Clin Psychol Sci* 2017; 5: 683–697.
24. Epstein LH, Paluch RA, Biondolillo MJ et al. Effects of 6-month episodic future thinking training on delay discounting, weight loss and hba1c changes in individuals with prediabetes. *J Behav Med* 2022; 45: 227–239.
25. O'Neill J, Daniel TO and Epstein LH. Episodic future thinking reduces eating in a food court. *Eat Behav* 2016; 20: 9–13.
26. Chiou WB and Wu WH. Episodic future thinking involving the nonsmoking self can induce lower discounting and cigarette consumption. *J Stud Alcohol Drugs* 2017; 78: 106–112.
27. Collado A and Stokes A. Imagining the future can shape the present: A systematic review of the impact of episodic future thinking on substance use outcomes. *Psychol Addict Behav* 2023; 38: 134–152.
28. O'Donnell S, Daniel TO and Epstein LH. Does goal relevant episodic future thinking amplify the effect on delay discounting?. *Conscious Cogn* 2017; 51: 10–16.
29. Athamneh LN, Stein MD, Lin EH et al. Setting a goal could help you control: Comparing the effect of health goal versus general episodic future thinking on health behaviors among cigarette smokers and obese individuals. *Exp Clin Psychopharmacol* 2021; 29: 59.
30. Schacter DL, Benoit RG and Szpunar KK. Episodic future thinking: Mechanisms and functions. *Curr Opin Behav Sci* 2017; 17: 41–50.
31. Bromberg U, Lobatcheva M and Peters J. Episodic future thinking reduces temporal discounting in healthy adolescents. *PLoS ONE* 2017; 12: e0188079.
32. Dassen FC, Jansen A, Nederkoorn C et al. Focus on the future: Episodic future thinking reduces discount rate and snacking. *Appetite* 2016; 96: 327–332.
33. Hollis-Hansen K, Seidman J, O'Donnell S et al. Episodic future thinking and grocery shopping online. *Appetite* 2019; 133: 1–9.
34. Stein JS, Tegge AN, Turner JK et al. Episodic future thinking reduces delay discounting and cigarette demand: an investigation of the good-subject effect. *J Behav Med* 2018; 41: 269–276.
35. Rinaldi G, Hijazi A and Haghparast-Bidgoli H. Cost and cost-effectiveness of mHealth interventions for the prevention and control of type 2 diabetes mellitus: A systematic review. *Diabetes Res Clin Pract* 2020; 162: 108084.
36. Sze YY, Daniel TO, Kilanowski CK et al. Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: A feasibility study. *JMIR Mhealth Uhealth* 2015; 3: e4603.
37. Leahey TM, Gorin AA, Wyckoff E et al. Episodic future thinking, delay discounting, and exercise during weight loss maintenance: The pace trial. *Health Psychol* 2020; 39: 796.
38. Meyerowitz-Katz G, Ravi S, Arnolda L et al. Rates of attrition and dropout in app-based interventions for chronic disease:

- systematic review and meta-analysis. *J Med Internet Res* 2020; 22: e20283.
39. Baumeister A, Muench F, Edan S et al. Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *J Med Internet Res* 2019; 21: e14567.
 40. Fuller-Tyszkiewicz M, Richardson B, Lewis V et al. A randomized trial exploring mindfulness and gratitude exercises as eHealth-based micro-interventions for improving body satisfaction. *Comput Human Behav* 2019; 95: 58–65.
 41. Eyles H, Jull A, Dobson R et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. *Curr Nutr Rep* 2016; 5: 160–167.
 42. Xu X, Griva K, Koh M et al. Creating a smartphone app for caregivers of children with atopic dermatitis with caregivers, health care professionals, and digital health experts: participatory co-design. *JMIR Mhealth Uhealth* 2020; 8: e16898.
 43. Baumeister A, Fleming T, Schueller SM et al. Digital micro interventions for behavioral and mental health gains: Core components and conceptualization of digital micro intervention care. *J Med Internet Res* 2020; 22: e20631.
 44. Persson DR, Yoo S, Bardram JE et al. Episodic future thinking as digital micro-interventions. In Kurosu M et al. (eds) *HCI International 2023 - Late breaking papers. HCII 2023. Lecture notes in computer science*, vol. 14054. Cham: Springer, pp. 213–237.
 45. Howe E, Suh J, Bin Morshed M et al. Design of digital workplace stress-reduction intervention systems: Effects of intervention type and timing. In *CHI conference on human factors in computing systems*. April 30–May 5 2022. New Orleans, LA, USA. Association for Computing Machinery, New York, NY, USA, Article 327, pp. 1–16.
 46. Meinschmidt G, Lee JH, Stalujanis E et al. Smartphone-based psychotherapeutic micro-interventions to improve mood in a real-world setting. *Front Psychol* 2016; 7: 1112.
 47. Klasnja P, Consolvo S and Pratt W. How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI conference on human factors in computing systems*. May 7–12 2011. Vancouver, BC, Canada. Association for Computing Machinery, New York, NY, USA, pp. 3063–3072.
 48. Bardram JE. The carp mobile sensing framework—a cross-platform, reactive, programming framework and runtime environment for digital phenotyping. *arXiv preprint arXiv:200611904* 2020.
 49. Chan CK and Cameron LD. Promoting physical activity with goal-oriented mental imagery: a randomized controlled trial. *J Behav Med* 2012; 35: 347–363.
 50. Levesque CS, Williams GC, Elliot D et al. Validating the theoretical structure of the treatment self-regulation questionnaire (TSRQ) across three different health behaviors. *Health Educ Res* 2007; 22: 691–702.
 51. Richards DK, Pearson MR and Field CA. Further validation of the treatment self-regulation questionnaire for assessing motivations for responsible drinking: A test of self-determination theory. *Exp Clin Psychopharmacol* 2021; 29: 679.
 52. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *Int J Hum Comput Interact* 2002; 14: 463–488.
 53. Johnson MW and Bickel WK. An algorithm for identifying non-systematic delay-discounting data. *Exp Clin Psychopharmacol* 2008; 16: 264.
 54. Twisk J, Bosman L, Hoekstra T et al. Different ways to estimate treatment effects in randomised controlled trials. *Contemp Clin Trials Commun* 2018; 10: 80–85.
 55. Koller M. *Robust estimation of linear mixed models*. PhD Thesis, ETH Zurich, 2013.
 56. Koller M. robustlmm: an r package for robust estimation of linear mixed-effects models. *J Stat Softw* 2016; 75: 1–24.
 57. Van Buuren S and Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in r. *J Stat Softw* 2011; 45: 1–67.
 58. Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; :–.
 59. Voss AT, Jorgensen MK and Murphy JG. Episodic future thinking as a brief alcohol intervention for heavy drinking college students: A pilot feasibility study. *Exp Clin Psychopharmacol* 2022; 30: 313.
 60. Vo V, Auroy L, Sarradon-Eck A et al. Patients’ perceptions of mHealth apps: Meta-ethnographic review of qualitative studies. *JMIR Mhealth Uhealth* 2019; 7: e13817.
 61. Schmaderer M, Miller JN, Mollard E et al. Experiences of using a self-management mobile app among individuals with heart failure: qualitative study. *JMIR Nurs* 2021; 4: e28139.
 62. Paredes P, Gilad-Bachrach R, Czerwinski M et al. Poptherapy: Coping with stress through pop-culture. In *Proceedings of the 8th international conference on pervasive computing technologies for healthcare (PervasiveHealth '14)*. May 20–23 2014. Oldenburg, Germany. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, pp. 109–117.
 63. Van Cappellen P, Catalino LI and Fredrickson BL. A new micro-intervention to increase the enjoyment and continued practice of meditation. *Emotion* 2020; 20: 1332.
 64. Conway N, Campbell I, Forbes P et al. mhealth applications for diabetes: user preference and implications for app development. *Health Inf J* 2016; 22: 1111–1120.
 65. de Witte M, Knapen A, Stams GJ et al. Development of a music therapy micro-intervention for stress reduction. *Arts Psychother* 2022; 77: 101872.
 66. Alexander B, Karakas K, Kohout C et al. A behavioral sensing system that promotes positive lifestyle changes and improves metabolic control among adults with type 2 diabetes. In *2017 systems and information engineering design symposium (SIEDS)*. Charlottesville, VA, USA, April 27–28 2017. IEEE, New York, USA, pp. 283–288.
 67. Jones A, Bardram JE, Bækgaard P et al. Integrated personalized diabetes management goes Europe: A multi-disciplinary approach to innovating type 2 diabetes care in Europe. *Prim Care Diabetes* 2021; 15: 360–364.
 68. Bickel WK, Jarmolowicz DP, Mueller ET et al. Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence. *Pharmacol Ther* 2012; 134(3): 287–297.
-